Quantifying into *wh*-dependencies:
Composing multi-*wh* questions and questions with quantifiers

Yimei Xiang, Rutgers University

Abstract *Wh*-questions with a quantificational subject have readings that seemingly involve quantification-into questions (called QɪQ for short). This paper argues to unify the derivation of QɪQ-readings and distinguish these readings from pair-list readings of multi-*wh* questions. I propose that QɪQ-questions and pair-list multi-*wh* questions both involve a *wh*-dependency relation, namely, that the trace of the subject-quantifier/*wh* stands in an anaphoric relation with the trace of the object-*wh*. In particular, in a pair-list multi-*wh* question, the subject-*wh* quantifies into an identity condition with respect to this dependency; in a QɪQ-question, the subject-quantifier quantifies-into a predication condition with respect to this dependency. This subtle difference yields the contrast with respect to domain exhaustivity. I further argue that the seeming QɪQ-effect in questions with quantifiers is derived by extracting a minimal proposition set that satisfies a quantificational predication condition. The possible values of this minimal set determine whether QɪQ-readings are available and whether a question admits a pair-list answer and/or a choice answer.

Keywords Questions, quantifiers, multi-*wh*, pair-list, functionality, uniqueness, domain exhaustivity, QV, categorial approaches, compositionality

1. Introduction

Questions with a subject universal quantifier (called ∀-questions for short henceforth) are ambiguous between individual readings, functional readings, and pair-list readings (Engdahl 1980, 1986). As exemplified in (1), the three readings expect answers naming an atomic movie, a Skolem function to atomic movies, and a list of boy-movie pairs, respectively.

(1) Which movie did every/each boy watch?

a. **Individual reading**
'For which movie y is s.t. every boy watch y?' '*Spiderman*.'

b. **Functional reading**
'For which function $\mathbf{f}$ to atomic movies is s.t. every-boy$_i$ x watched $\mathbf{f}(x)$?'
'His$_i$ favorite superhero movie.'

c. **Pair-list reading**
'For every boy x, [tell me] which movie did x watch?'
'Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.'

There are two general ways to think about the nature of the pair-list reading (1c). One way regards this reading as involving quantification-into questions (abbreviated as 'QɪQ' henceforth) (Groenendijk and Stokhof 1984; Chierchia 1993; among others). An informal paraphrase for QɪQ-readings is given in (2), where 'Det' stands for a determiner.

(2) Which movie did Det-boy(s) watch? (QɪQ-reading)
≈ 'For Det-boy(s), [you tell me]/[I ask you] which movie did they watch?'

For questions with an existential indefinite (henceforth called ∃-questions), their QɪQ-readings have a choice flavor (Groenendijk and Stokhof 1984). For example, the choice reading (3b) asks to choose one/two of the relevant boys and specify the unique movie he/they watched.[1] In contrast, questions with a negative quantifier (henceforth called NO-questions) do not have QIQ-readings. For example, (4) cannot be responded by silence.

(3) Which movie did one/two of the boys watch?

a. **Individual reading**
'For which movie y is s.t. one/two of the boys watched y?' '*Ironman*.'

b. **Choice reading**
'For one/two of the boys, [you tell me] which movie did he/they watch?'
'Andy watched *Ironman*.'/ 'Billy and Clark watched *Spiderman*.'

(4) Which movie did {no boy, none of the boys} watch?

a. **Individual reading**
'For which movie y is s.t. no boy watched y?' '*Revengers*.'

b. **Functional reading**
'For which function $\mathbf{f}$ to atomic movies is s.t. no boy x watched $\mathbf{f}(x)$?'
'The movie recommended by their grandfather.'

c. # **QIQ-reading**
'For no boy, [you tell me] which movie did they watch?' [Slience]

The other way to group the aforementioned types of complex questions is to treat questions with pair-list readings uniformly. Similar to the ∀-question (1), the multi-*wh* question (5) also has a reading that requests to specify a list of boy-movie pairs. Accounts adopting this line of thinking either use the same LF to compose the ∀-question (1) and the corresponding multi-*wh* questions (5) (Engdahl 1980, 1986; Dayal 1996, 2017) or assign these two questions with the same root denotation (Fox 2012a,b).

(5) Which boy watched which movie?

a. **Single-pair reading**
'Which unique boy-x-to-movie-y pair is such that x watched y?'
Andy watched *Spiderman*.

b. **Pair-list reading**
'What boy-x-to-movie-y pairs are such that x watched only y?'
'Andy watched *Ironman*, Billy watched *Spiderman*, Clark watched *Hulk*.'

In sum, it is controversial whether we should treat questions with QIQ-readings (QIQ-questions henceforth) uniformly or questions with pair-list readings (pair-list questions henceforth) uniformly. This paper argues for the former option. On the one hand, pair-list readings of ∀-questions and multi-*wh* questions differ with respect to domain exhaustivity (Sect. 2.1). This contrast suggests that these two types of pair-list questions have different root denotations and procedures of composition.

[1] Functional readings are marginally acceptable for ∃-questions. For example, the fragment functional answer (ia) sounds under-informative. The boy who watched the movie has to be specified, as in (ib). I leave this puzzle open.

(i) Which movie did one of the boys watch?
(*w*: *Among the relevant boys, only Andy watched a movie, which was his favorite superhero movie — Ironman.*)

a. ?? His favorite superhero movie.

b. Andy watched his favorite superhero movie.

On the other hand, the similarities between these two types of questions in form and meaning also suggest that their composition procedures should not be drastically different.

I propose that QiQ-questions and pair-list multi-*wh* questions both involve *wh*-dependencies, namely, the trace of the subject quantifier/*wh* stands in an anaphoric/functional relation with the trace of the object-*wh*. The core assumptions of this analysis are illustrated in (6). The *wh*-dependency is realized by assigning an additional index (i.e., the index of the trace of the subject-*wh*/quantifier) to the trace of the object-*wh* (Sect. 4.1.1). I further assume that in (6a) the subject-quantifier quantifies into a predication (PRED) condition with respect to this dependency relation, and that in (6b) the subject-*wh* quantifies into an identity (IDENT) condition with respect to this dependency relation. As we will see in Sect. 6, the differences between these two quantifying-in operations can naturally explain the contrast between ∀-questions and multi-*wh* questions with respect to domain exhaustivity.

(6) **A general schema of composing complex questions**

a. Which movie did DET-boy(s) watch? (QIQ-reading)
... [which-movie$_j$ ... DET-boy(s)$_i$ [$_{\text{PRED}}$ ... [$_{\text{VP}}$ t_i watched t_j^i]]]

b. Which boy watched which movie? (Pair-list reading)
... [which-movie$_j$ ... which-boy$_i$ [$_{\text{IDENT}}$ ... [$_{\text{VP}}$ t_i watched t_j^i]]]

The rest of this paper is organized as follows. Section 2 presents evidence against the view of unifying pair-list (∀- and multi-*wh*) questions as well as evidence for the view of composing QIQ-questions uniformly. Section 3 lays out the technical challenges in composing QIQ-questions and the related semantic phenomena that this paper aims to account for. The phenomena include the contrast between ∀- and multi-*wh* questions with respect to domain exhaustivity, the point-wise uniqueness effects in pair-list questions with a singular-marked *wh*-object, the limited distribution of pair-list in matrix QIQ-questions, and the quantificational variability (QV) effects in embeddings of pair-list questions. Section 4 reviews two influential approaches to composing pair-list questions, including the functionality-based approach of Dayal 1996 and the family-of-question approach of Fox 2012a,b. Section 5 introduces a hybrid categorial approach to question composition (Xiang 2016, 2020), which I use as a general framework of composing questions. The core analysis in this paper is independent from this framework, but this framework allows to derive QV effects without assuming a non-flat semantics. Section 6 puts forward my central analysis of composing pair-list multi-*wh* questions and QIQ-questions. The denotations and the composition procedures of these two types of questions will be presented in tandem. Section 7 accounts for the QV effects in embeddings of pair-list questions. Section 8 concludes.

2. Arguments for unifying the derivation of QIQ-readings

This section argues that pair-list ∀-questions should be composed uniformly as other QIQ-questions, not as pair-list multi-*wh* questions. On the one hand, when having pair-list readings, ∀-questions are subject to a domain exhaustivity condition, while their multi-*wh* counterparts are not (Sect. 2.1). This contrast suggests that these two types of questions should be interpreted and derived differently. On the other hand, evidence from syntactic distributions suggests that QIQ-questions have a uniform syntax — in these questions, QIQ-readings exhibit the same subject-object/adjunct asymmetry, and moreover, the distributional pattern of QIQ-readings is preserved in questions where the subject is a coordination of quantifiers (Sect. 2.2).

2.1. A contrast in domain exhaustivity

It is commonly thought that pair-list readings of multi-*wh* questions and ∀-questions are both subject to **domain exhaustivity** (Dayal 1996, 2002; among others). For a question with a *wh*/∀-subject and a *wh*-object, the domain exhaustivity condition says that every member of the set quantified over by the *wh*/∀-subject must be paired with a member of the set quantified over by the *wh*-object. For instance, in (1) and (5), repeated below, domain exhaustivity requires that <u>every</u> boy watched a (possibly different) movie. Moreover, since the object-*wh* is singular-marked (viz., the *wh*-complement is singular), the two questions are also subject to **point-wise uniqueness**, which says that each boy watched <u>at most one</u> movie.

(7) a. Which movie did every/each boy watch?
b. Which boy watched which movie?

While the point-wise uniqueness effect is easy to attest, the domain exhaustivity effect is quite obscure. For example, in the multi-*wh* question (7a), it is unclear which set of boys is quantified over by the subject-*wh*; domain exhaustivity would be trivial if this quantification domain consists of only the boys who did watch a movie. To remove this confound, Fox (2012a) uses the pair of examples in (8), where the quantification domain of each *wh*-phrase is explicitly specified. Fox claims that (8b) rejects a pair-list reading (in contrast to (8a)), arguing that this reading is rejected because the domain exhaustivity condition presupposed in a pair-list reading is contextually infelicitous — pairing four kids with three chairs yields that there will be multiple kids sitting on the same chair.

(8) a. Guess which one of the **three** kids will sit on which one of the **four** chairs.
b. Guess which one of the **four** kids will sit on which one of the **three** chairs.

In contrast to the dominant view, I argue that pair-list multi-*wh* questions are <u>not</u> subject to domain exhaustivity. First, pair-list multi-*wh* questions can be felicitously used in contexts where domain exhaustivity is violated. In (9), the sentence copied from (8b) is fully acceptable and must be interpreted with a pair-list reading.

(9) (w: *Four kids are playing Musical Chairs and are competing for three chairs.*)
"Guess which one of the **four** kids will sit on which one of the **three** chairs."
$\not\rightsquigarrow$ *Each of the four kids will sit on one of the three chairs.*

The game rules of Musical Chairs yield two conditions: (i) one of the four kids will not sit on any of the three chairs, and (ii) the rest three kids each will sit on a different chair. Condition (ii) ensures that the embedded multi-*wh* question has a pair-list reading, not a single-pair reading. Condition (i) contradicts the domain exhaustivity inference that each of the kids will sit on one of the chairs. If pair-list multi-*wh* questions were subject to domain exhaustivity, (9) would suffer a presupposition failure and would be infelicitous in the given context, contra fact.

Second, in contrast to their multi-*wh* counterparts, pair-list ∀-questions cannot be felicitously used in contexts where domain exhaustivity is violated. In the context in (10), the quantification domain of the subject-*wh*/quantifier is greatly larger than that of the object-*wh*. The multi-*wh* question (10a) is fully acceptable, but the ∀-question (10b) is not: (10b) presupposes that each candidate will get one of the jobs, contra context.

(10) (w: **100** *candidates are competing for* **three** *job openings.*)

a. ✓ "Guess which candidate will get which job."

b. # "Guess which job will every candidate get."

One might suggest that the domain exhaustivity condition of a multi-*wh* question can be associated with any of the *wh*-phrases, including also the object-*wh*. For example, in (9) and (10), it could be the case that domain exhaustivity requires every chair and every job to be taken by a kid and a candidate, respectively. However, this possibility is also ruled out: a pair-list multi-*wh* question can be uttered in a context where neither type of domain exhaustivity is satisfied. For example, the sentence (11) is felicitous, and it does not imply domain exhaustivity relative to boys or to girls.

(11) (*w*: *Four boys and four girls will form four boy-girl pairs to perform in a dance competition, but only* **two** *of the pairs will get into the final round.*)
"Guess which one of the **four** boys will dance with which one of the **four** girls in the final round."
$\not\rightsquigarrow$ *Each of the four boys will dance with one of the four girls in the final round.*
$\not\rightsquigarrow$ *Each of the four girls will dance with one of the four boys in the final round.*

In conclusion, pair-list readings of ∀-questions are subject to domain exhaustivity, while pair-list readings of multi-*wh* questions are not. This contrast suggests that these two pair-list questions should be interpreted and composed differently.

2.2. Uniform distribution of QɪQ-readings

The distribution of QɪQ-readings uniformly exhibits a subject-object/adjunct asymmetry (May 1985, 1988; Chierchia 1991, 1993). As seen in (12) and (13), pair-list readings and choice readings are available if the non-*wh* quantifier serves as the subject while the *wh*-phrase serves as the object, and otherwise are unavailable. In (12b), the uniqueness inference triggered by the singular-marked *wh*-subject has to be interpreted with wide scope relative to the object universal quantifier. As for the ∃-questions in (13), despite that (13b) marginally admits a choice reading, (13a) is much more preferable if the questioner seeks for a choice answer.[2] The subject-adjunct asymmetry is analogous, as illustrated in (14) and (15). Thus, unless there is compelling evidence to suggest otherwise, it is appealing to assume that QɪQ-readings are derived uniformly.

(12) (*w*: *Ten students made votes for three candidates. Each student voted for only one candidate. The questioner wants to know all of the student-candidate pairs*)

a. Which candidate did every student vote for? (✓Pair-list)

b. # Which student voted for every candidate? (✗Pair-list)
$\rightsquigarrow$ *Exactly one of the students voted for every candidate.*

(13) (*w*: *Ten students made votes for three candidates. Each student voted for only one candidate. The questioner is only interested in knowing one of the student-candidate pairs.*)

a. Which candidate did one of the students vote for? (✓Choice)
Andy voted for the first candidate.

b. ? Which student voted for one of the candidates? (?Choice)

(14) (*w*: *Each driver refueled at a nearby station exactly once.*)

[2]The reason why (13b) and (15) marginally admit choice readings might be that existential indefinites have more ways to take scope than universal quantifiers, such as through choice functions.

a. At which station did every driver refuel? (✓Pair-list)

b. # Which driver refueled at every gas station? (✗Pair-list)

(15) (*w*: *Each driver refueled at a nearby station exactly once.*)

a. At which station did [one of the drivers] refuel? (✓Choice)

b. ? Which driver refueled at [one of the nearby stations]? (?Choice)

The view of unifying QiQ-readings is further supported by the interpretations of questions with a coordination of quantifiers. In (16a) where the subject is a conjunction of a universal quantifier and an existential indefinite, the pair-list reading associated with the universal quantifier and the choice reading associated with the existential indefinite are both preserved. This question can be understood as requesting to specify all boy-watch-movie pairs and one girl-watch-movie pair. In contrast, since negative quantifiers do not license QiQ-readings (recall (4)), coordinating a universal/existential quantifier with a negative quantifier blocks the QiQ-reading. For example, (16b) cannot be read as requesting to list all boy-watch-movie pairs and not to list any teacher-watch-movie pairs.

(16) a. Which movie did [each of the boys and one of the girls] watch? (✓QiQ)

b. Which movie did [each of the boys and none of the teachers] watch? (✗QiQ)

c. Which movie did [one of the girls and none of the teachers] watch? (✗QiQ)

3. Challenges and goals

Section 2 has laid out two goals for this paper: (i) to derive the QiQ-readings of questions with quantifiers uniformly, and (ii) to compose pair-list multi-*wh* questions in tandem with pair-list ∀-questions while explaining their contrast with respect to domain exhaustivity. However, it is not easy to achieve both goals. This section discusses the technical challenges that need to be overcome and the related semantic effects that need to be accounted for.[3]

First, for most frameworks of question semantics, the structure in (17) is ill-formed. The generalized quantifier 'Det-boy' take arguments of type $\langle e, t\rangle$ and can only quantify into a t-type expression. However, the contained open question 'which movie did x watch' is not of type t; instead, it has been treated, for example, as a set of propositions (of type $\langle st, t\rangle$) as in Hamblin-Karttunen Semantics, or as a one-place predicate/property (of type $\langle e, t\rangle$ or $\langle e, st\rangle$) as in categorial approaches.

(17) Which movie did Det-boy watch?
*[Det-boy λx_e [which movie did x watch]]

There are two general strategies to solve this type-mismatch problem. One is to extract the domain of quantification of the subject-quantifier via a type-shifting operation (Groenendijk and Stokhof 1984; Chierchia 1993; Dayal 1996, 2017; among others). For example, Dayal extracts the quantification domain of a universal quantifier as extracting the unique minimal witness set of the quantifier. This strategy is feasible in principle but a bit *ad hoc* (see Sect. 4.1.2 and footnote 10).

The other strategy is to create a t-type node in the LF which the quantifier can quantify into. For example, in Partition Semantics (Groenendijk and Stokhof 1984) which defines the root denotation of a question as a partition of possible worlds, the formation of a partition involves a t-type

[3] This paper does not attempt to explain effects that more likely to be related to syntax in nature, such as the superiority effects and constraints of extractions/movements. See Kotek 2014, 2019 and the references therein for detailed discussions.

node expressing an identity condition. Alternatively, Karttunen (1977) and Krifka (2001) reduce quantification-into matrix questions into quantification-into question-embeddings. The two analyses based on partitions and question-embeddings overcome the type-mismatch problem but bring up other problems (for reviews, see Appendices). Instead, my proposal will follow Fox (2012b) in assuming that the root of a QiQ-question contains a t-type node that expresses a predication condition (Sect. 4.2 and 6.3).[4]

Second, pair-list readings have a limited distribution in matrix QiQ-questions. In matrix questions, only subject *each*/*every*-phrases can license pair-list readings. For example, in the ∃-question (18) which has a numeral-modified indefinite *two of the students*, the seeming pair-list answer (18a) which distributes over two chosen students is actually an over-informative specification of a cumulative choice answer (18b) (Moltmann and Szabolcsi 1994; Szabolcsi 1997a). Questions with a plural *the*-phrase like (19) are analogous (Srivastav 1991; Krifka 1991).

(18) Who did two of the students vote for?
  a. Andy voted for Mary, and Billy voted for Jill.
  b. Andy and Billy voted for Mary and Jill. In particular, Andy voted for Mary, and Billy voted for Jill.

(19) Who did the students vote for?

The confound from cumulative answers can be removed by replacing the number-neutral word *who* with a singular-marked *wh*-phrase, which triggers a uniqueness presupposition. In the following set of matrix questions, distributivity above uniqueness is possible only in (20a-b), where the subject quantifier is distributive in lexicon. In other cases, for example, the choice reading of the ∃2-question (20d) presupposes that two of the students voted for the same candidate and only this candidate, conflicting with the context.

(20) I know that every student voted for a different candidate. Which candidate did ...
  a. ... **every** student vote for? ($\forall \gg \iota$)
  b. ... {**each** student, **each** of the students} vote for? (EACH $\gg \iota$)
  c. # ... **all**/**most** of the students vote for? (ALL/MOST $\gg$ ~~EACH~~ $\gg \iota$)
  d. # ... **two** of the students vote for? ($\exists 2 \gg$ ~~EACH~~ $\gg \iota$)
  e. # ... **two or more** students vote for? ($\exists 2+ \gg$ ~~EACH~~ $\gg \iota$)
  f. # ... **the** students vote for? (THE-$\text{NP}_{\text{PL}} \gg$ ~~EACH~~ $\gg \iota$)

To account for the limited distribution of pair-list readings in matrix questions, many works on composing complex questions propose to derive pair-list readings in a way that crashes in questions with a non-universal quantifier (e.g., Dayal 1996 and Fox 2012b; see Sect. 4 for details.) This strategy, however, comes with an expense of failing to account for choice readings of ∃-questions. In contrast, I argue that a subject-quantifier licenses pair-list readings only if this quantifier is lexically distributive and scopally productive. In this view, the limited distribution of pair-list naturally follows from the independently observed contrasts between distributive-universal quantifiers and the other quantifiers with respect to lexical distributivity and scoping (Szabolcsi 1997b; Beghelli and Stowell 1997; for details, see Sect. 6.3.2 and 6.3.4).

[4]Other than these two general strategies, Inquisitive Semantics also exempts from this type-mismatch problem because it defines declaratives and interrogatives uniformly as a set of sets of propositions (of type $\langle stt, t\rangle$) and generalized quantifiers as functions of type $\langle\langle e, stt\rangle, t\rangle$. To my knowledge, this idea has not been explored extensively. For a possible direction, see Ciardelli and Roelofsen (2018: Sect. 4.3.3).

Third, there are several semantic effects robustly observed with QIQ-questions and/or pair-list *wh*-questions. Section 2.1 has discussed two effects, including the **uniqueness** effect triggered by the singular-marked object-*wh*, as seen in all the sentences in (21), and the **domain exhaustivity** effect observed only in ∀-questions, as seen in (21a). These effects were not extensively considered until Srivastav 1991/Dayal 1996.

(21) a. Which movie did every/each boy watch?
$\rightsquigarrow$ *For every boy x, x watched exactly one movie.*
b. Which boy watched which movie?
$\rightsquigarrow$ *For every boy x such that x watched any movie, x watched exactly one movie.*
c. Which movie did one/two of the boys watch?
$\rightsquigarrow$ *For some x such that x is one/two of the boys, x watched exactly one movie.*

Moreover, embeddings of pair-list questions are subject to **quantificational variability** (QV) effects. As first observed by Berman (1991), question-embeddings modified by a quantificational adverbial (e.g., *mostly, partly, for the most part, in part*) commonly have a QV inference. As illustrated in (22) and (23), in paraphrasing such an inference, the quantification domain of the matrix quantity adverbial *mostly* can be thought of as (a) a set of propositions (Lahiri 1991, 2002; Cremers 2016), (b) a set of sub-questions (Beck and Sharvit 2002), or (c) a set of individuals or pairs (Xiang 2016, 2019b, 2020; Cremers 2018). This effect casts challenges to accounts such as Dayal 1996 which analyzes pair-list questions with a flat semantics (Sect. 4.1.2).

(22) Jill mostly knows [which students left].
a. $\rightsquigarrow$ *For most p: p is a true proposition of the form ⌜student-x left⌝, Jill knows p.*
b. $\rightsquigarrow$ *For most Q: Q is a question of the form ⌜whether student-x left⌝, Jill knows Q.*
c. $\rightsquigarrow$ *For most x: x is an atomic student and x left, Jill knows that x left.*

(23) Jill mostly knows [$_{\text{PAIR-LIST}}$ { which movie every boy watched / which boy watched which movie }].
a. $\rightsquigarrow$ *For most p: p is a true proposition of the form ⌜boy-x watched movie-y⌝, Jill knows p.*
b. $\rightsquigarrow$ *For most Q: Q is a question of the form ⌜which movie boy-x watched⌝, Jill knows Q.*
c. $\rightsquigarrow$ *For most $\langle x, y\rangle$: x is an atomic boy and y is an atomic movie and x watched y, Jill knows that x watched y.*

4. Two general approaches to composing complex questions

There is a rich literature on composing pair-list multi-*wh* questions and questions with quantifiers. This section reviews two lines of approaches that have tackled both types of questions, including the **functionality-based approaches** which assume that these complex questions involve *wh*-dependencies, and the **family-of-question approaches** which define each of such questions as a family of sub-questions.[5] I will especially focus on two influential accounts, namely, Dayal (1996, 2017) and Fox (2012a,b), because they successfully predict the domain exhaustivity and point-wise uniqueness effects in singular-marked ∀-questions, and because my analysis will take ingredients from these two accounts. For more extensive reviews, see the Appendices as well as Xiang 2016: chapter 5 and 6, Dayal 2017: chapter 4, and Ciardelli and Roelofsen 2018.

[5] The core assumptions of these two approaches are compatible with each other. For example, Chierchia (1993) assumes *wh*-dependency while defining a QIQ-question as a family of questions. See details in footnote 10.

4.1. Function-based approaches

Functional readings of questions with quantifiers exhibit a clear functional dependency relation between the subject-quantifier and the object-*wh*, called "*wh*-dependency". In example (1b), repeated below, the answer involves a pronoun interpreted as being bound by the subject-quantifier in the question.

(24) Which movie did every-boy$_i$ watch?
His$_i$ favorite superhero movie.

As for pair-list readings of questions, functionality-based approaches assume that ∀-questions and multi-*wh* questions with pair-list readings also involve a *wh*-dependency relation between the higher ∀/*wh*-phrase and the lower *wh*-phrase. In this view, for example, the pair-list answer (25a) specifies the graph of a Skolem function from the set that the higher ∀/*wh*-phrase ranges over to the set that the lower *wh*-phrase ranges over, as in (25b).

(25) Which movie did every boy watch?/ Which boy watched which movie?

a. Andy watched *Ironman*,
Billy watched *Spiderman*,
Clark watched *Hulk*.

b. $\mathbf{f} = \begin{bmatrix} a & \rightarrow & i \\ b & \rightarrow & s \\ c & \rightarrow & h \end{bmatrix}$

The functionality-based analysis was originally proposed only for ∀-questions (Engdahl 1980, 1986; Chierchia 1993), especially to account for the similar subject-object/adjunct asymmetry in their functional readings and their pair-list readings. This asymmetry is illustrated by the contrast between (26) and (27) (see also Sect. 2.2): functional readings and pair-list readings are available only if the universal quantifier is higher than the *wh*-phrase in the syntactic structure. Assuming functionality, one can explain this asymmetry in terms of Weak Crossover Violations or the Left-ness Constraint in binding and functionality (Chierchia 1993; Jacobson 1994; Williams 1994).

(26) Which woman did every boy invite? (✓Individual, ✓Functional, ✓Pair-list)
a. Anna.
b. His mother.
c. Andy invited Mary, Billy invited Susi, Clark invited Jill.

(27) Which woman invited every boy? (✓Individual, ✗Functional, ✗Pair-list)
a. Anna.
b. # His mother. (Intended: 'Every-boy$_i$ was invited by his$_i$ mother.')
c. # Andy invited Mary, Billy invited Susi, Clark invited Jill.

Further, Dayal (1996, 2017) extends the idea of functionality to pair-list multi-*wh* questions. She points out that the corresponding relations expressed by pair-list answers are skolem functions — the correspondence can be one-to-one or many-to-one, but not one-to-many, as witnessed in (28). See also Caponigro and Fălăuş (To appear) for an extension of this approach to multi-*wh* free relatives in Romanian.

(28) Which student talked to which professor? (Dayal 2017: 96)
a. Alice talked to Professor Carl, and Bill talked to Professor Dan.
b. Alice and Bill both talked to Professor Carl.

c. # Alice talked to Professors Carl and Dan.

This paper does not take a position on whether the subject-object/adjunct asymmetry and the unavailability of one-to-many relations should be explained in terms of constraints in functionality. However, in section 6, proposing a new compositional analysis, I will show that *wh*-dependency is independently needed to account for the contrast between multi-*wh* questions and ∀-questions with respect to domain exhaustivity.

4.1.1. *Wh*-dependency in basic functional questions

In the current dominant analysis, *wh*-dependencies in functional questions are derived by assuming a complex *wh*-trace (Groenendijk and Stokhof 1984; Chierchia 1993; among others).[6] The tree diagram in (29) illustrates the LF schema for a ∀-question with a functional reading.[7] In this LF, the *wh*-trace t_i^j carries two indices, including:

(i) a functional index i, which is interpreted as an intensional functional variable $\mathbf{f}$ (of type $\langle s, ee\rangle$) and is bound by the fronted object-*wh which movie*;

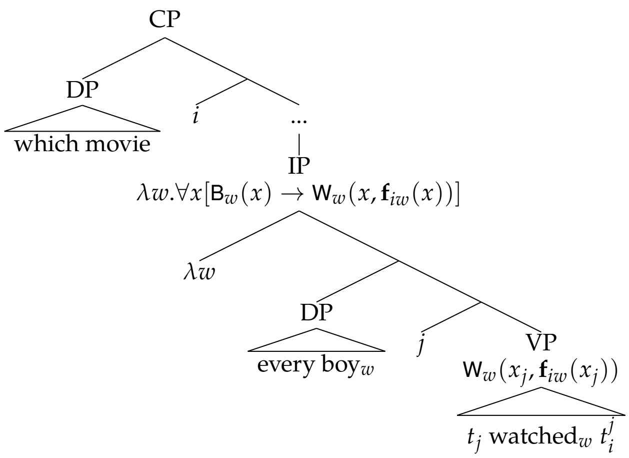
(ii) an argument index j, which is interpreted as an individual variable x (of type e) and is bound by the subject-quantifier *every boy*.

With the above binding relations, the IP is interpreted as an open proposition expressing a quantificational functional dependency condition, read as 'every boy x watched $\mathbf{f}(x)$'. The details of composition above IP are omitted for now because they vary by the framework of question composition. For example, in Hamblin-Karttunen Semantics, the yielded root denotation of this question is a set of propositions of the form ⌜*every boy x watched* $\mathbf{f}(x)$⌝ where $\mathbf{f}$ is an intensional Skolem function to atomic movies (viz., $\forall w[\text{Ran}(\mathbf{f}_w(x)) \subseteq \mathsf{M}_w]$, or equivalently, $\forall w \forall x \in \text{Dom}(\mathbf{f}_w)[\mathsf{M}_w(x)]$), as in (30a). In categorial approaches, the yielded denotation is a property/predicate of these intensional Skolem functions, as in (30b).

(29) Which movie did every boy watch? (Functional reading)

[6]Other than the complex trace approach, the variable-free approach of Jacobson (1999) does not use indices/variables at all. Instead, functional dependency is derived by a locally applied **z**-rule which can close off the anaphoric dependency between the arguments of a predicate. The *wh*-trace is interpreted as an identity function over Skolem functions $\lambda f_{\langle e,e\rangle}.f$, and the abstraction λf is passed up to the entire question nucleus by the application of another type-shifting rule — the Geach (**g**)-rule. For ease of comparing with existing works on composing complex questions, this paper follows the complex functional trace approach. For an attempt of using the variable-free approach to compose complex questions, see Xiang 2019b.

[7]Following Groenendijk and Stokhof (1984), I translate LF representations into the Two-sorted Type Theory (Ty2) of Gallin (1975). Compared with Montague's Intensional Logic, Ty2 is different in that it introduces s (the type of possible worlds) as a basic type (just like e and t), and in that it uses variables and constants of type s which can be thought of as denoting possible worlds. For example, the English common noun *boy* is translated into B_w in Ty2, where B is a property of type $\langle s, et\rangle$ and w a variable of type s. With these assumptions, Ty2 can make direct reference to worlds and allows quantification and abstraction over world variables.



(30) a. Question denotation in Hamblin-Karttunen Semantics
$[\![\text{CP}]\!] = \{\lambda w.\forall x[\mathsf{B}_w(x) \rightarrow \mathsf{W}_w(x, \mathbf{f}_w(x))] \mid \forall w[\text{Ran}(\mathbf{f}_w(x)) \subseteq \mathsf{M}_w]\}$

b. Question denotation in categorial approaches
$[\![\text{CP}]\!] = \lambda \mathbf{f}_{\langle s,ee \rangle} : \forall w[\text{Ran}(\mathbf{f}_w) \subseteq \mathsf{M}_w].\ \lambda w.\forall x[\mathsf{B}_w(x) \rightarrow \mathsf{W}_w(x, \mathbf{f}_w(x))]$

4.1.2. Dayal (1996, 2017) on composing pair-list questions

Dayal (1996, 2017) assumes that the two pair-list questions in (31) both denote a set of conjunctive propositions, and that each of the contained conjunctive propositions specifies a Skolem function $\mathbf{f}$ from the quantification domain of the $\forall$/*wh*-subject (i.e., $\mathsf{B}_@$) to the quantification domain of the *wh*-object (i.e., $\mathsf{M}_@$).[8] This denotation yields domain exhaustivity since the function $\mathbf{f}$ takes the set of atomic boys as its domain.

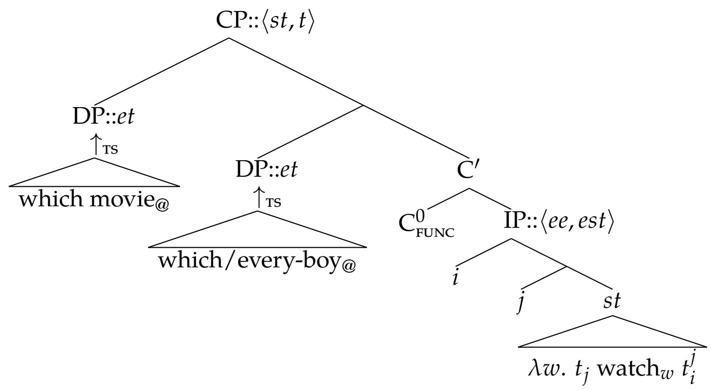
(31) Which movie did every boy watch?/ Which boy watched which movie?
(Context: *There are two relevant boys $b_1 b_2$ and two relevant movies $m_1 m_2$.*)

$$[\![Q_\forall]\!] = [\![Q_{\text{multi-wh}}]\!] = \{\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\} \mid \mathbf{f} \in [\mathsf{B}_@ \rightarrow \mathsf{M}_@]\}$$
$$= \left\{ \begin{array}{l} \lambda w.\mathsf{W}_w(b_1, m_1) \wedge \mathsf{W}_w(b_2, m_1) \\ \lambda w.\mathsf{W}_w(b_1, m_1) \wedge \mathsf{W}_w(b_2, m_2) \\ \lambda w.\mathsf{W}_w(b_1, m_2) \wedge \mathsf{W}_w(b_2, m_1) \\ \lambda w.\mathsf{W}_w(b_1, m_2) \wedge \mathsf{W}_w(b_2, m_2) \end{array} \right\}$$

Dayal assumes that both of the pair-list questions in (31) are composed via the LF (32). In this LF, both the subject-*wh*/quantifier and the object-*wh* are moved to the specifier of the projection of a functional C head C^0_{FUNC}.

[8] '@' stands for the actual world. For simplicity, here and henceforth, I assume that the extensions of the *wh*-complements are evaluated relative to the actual world.

(32)



(33) a. $[\![\text{IP}]\!] = \lambda \mathbf{f}_{\langle e,e\rangle} \lambda x_e \lambda w.\mathsf{W}_w(x, \mathbf{f}(x))$

b. $[\![\text{C}^0{}_{\text{FUNC}}]\!] = \lambda q_{\langle ee,est\rangle} \lambda D \lambda R \lambda p.\exists \mathbf{f} \in [D \to R][p = \bigcap \lambda p'.\exists x \in D[p' = q(\mathbf{f})(x)]]$
$= \lambda q_{\langle ee,est\rangle} \lambda D \lambda R.\{\bigcap\{q(\mathbf{f})(x) \mid x \in D\} \mid \mathbf{f} \in [D \to R]\}$

c. $[\![\text{C}']\!] = \lambda D \lambda R \lambda p.\{\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid x \in D\} \mid \mathbf{f} \in [D \to R]\}$

d. $[\![\text{CP}]\!] = \{\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid x \in \mathsf{B}_@\} \mid \mathbf{f} \in [\mathsf{B}_@ \to \mathsf{M}_@]\}$

The composition precedes in three steps. **First**, the trace of the *wh*-object that carries two indices — a functional index i interpreted as an $\langle e, e\rangle$-type variable $\mathbf{f}$, and an argument index j interpreted as an e-type variable x. The trace of the *wh*/∀-subject also carries the argument index j. Abstracting the two indices at the edge of IP yields a two-place property (of type $\langle ee, est\rangle$). As defined in (33a), this property maps a Skolem function $\mathbf{f}$ and an individual x to an open proposition that expresses a functional dependency relation between the subject and the object of *watch*. **Second**, as in (33b-c), the complex head C^0_{FUNC} introduces domain and range arguments for the Skolem function $\mathbf{f}$ and creates a graph for $\mathbf{f}$. For q being the denotation of IP, the graph of a Skolem function $\mathbf{f}$ yielded by q is the conjunction of propositions of the form $\ulcorner q(\mathbf{f})(x)\urcorner$ where x is in the domain of $\mathbf{f}$. **Last**, the sets that the ∀/*wh*-phrases range over are extracted by type-shifting operations (indicated by '$\uparrow_{\text{TS}}$') and are passed to fill the range and domain arguments of C^0_{FUNC}.[9] With this composition, the denotation of the question root (i.e., CP) is a set of conjunctive propositions, each of which names a Skolem function defined for the set that the *wh*/∀-subject ranges over. This domain condition gives rise to a domain exhaustivity effect.

Finally, to account for the uniqueness effects of singular-marked *wh*-phrases, Dayal defines an answerhood-operator that presupposes the existence of the strongest true answer. The strongest true answer to a question is the true proposition in the Hamblin set of this question that entails all the true propositions in this set.

(34) $\text{Ans}_{Dayal}(w)(Q) = \exists p[w \in p \in Q \wedge \forall q[w \in q \in Q \to p \subseteq q]].$
$\iota p[w \in p \in Q \wedge \forall q[w \in q \in Q \to p \subseteq q]]$

The ontology of individuals assumes that a singular noun denotes a set of atomic entities, while a plural noun denotes a set consisting of both atomic and sum entities (Sharvy 1980; Link 1983). If sums are defined in terms of part-hood relation, this ontology can be represented as in Figure 1. Letters *abc* each denotes an atomic boy. Lines indicate *part of* relations from bottom to top.

[9] Dayal (2017) discusses two ways to obtain the quantification domain of a *wh*-phrase. One way is to define a *wh*-phrase as an existential quantifier and extract out its quantification domain via the application of a Be-shifter (Partee 1986). The other way is to define a *wh*-phrase as a set of entities and derive its quantificational meaning via employing an ∃-shifter.

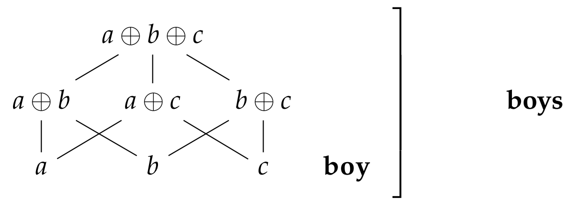


Figure 1: Ontology of individuals (Sharvy 1980; Link 1983)

Accordingly, the Hamblin set of a singular-marked *wh*-question (35a) includes only propositions naming an atomic boy, while the Hamblin set of the corresponding plural-marked question (35b) includes also propositions naming a sum of boys. In a discourse where both Andy and Bill watched *Hulk*, the true answers are given in (35a'-b'). Note that the set (35b') has a strongest proposition $\lambda w.\mathsf{W}_w(a \oplus b, h)$ but (35a') does not; therefore, employing Ans_{Dayal} in (35a) gives rise to a presupposition failure. To avoid this presupposition failure, the singular-marked question (35a) can only be felicitously uttered in a world where only one of the boys watched *Hulk*, which therefore explains its uniqueness requirement.

(35) (w: *Among the considered boys, only Andy and Billy watched Hulk.*)
  a. Which boy watched *Hulk*? a′. $\{\lambda w.\mathsf{W}_w(a,h), \lambda w.\mathsf{W}_w(b,h)\}$
  b. Which boys watched *Hulk*? b′. $\{\lambda w.\mathsf{W}_w(a,h), \lambda w.\mathsf{W}_w(b,h), \lambda w.\mathsf{W}_w(a \oplus b,h)\}$

In a pair-list question, if the object-*wh* is singular-marked, the presupposition of Ans_{Dayal} entails point-wise uniqueness. For example, if in w_1 the boy b_1 watched only m_1 but b_2 watched both $m_1 m_2$, then the top two propositions in the Hamblin set Q in (31) are both true in w_1. Since neither of the true propositions is stronger than the other, applying $\text{Ans}_{Dayal}(w_1)$ to Q yields a presupposition failure.

The account of Dayal successfully predicts domain exhaustivity and point-wise uniqueness effects in ∀-questions with a singular-marked *wh*-object. In this account, domain exhaustivity comes from the lexical meaning of C^0_{FUNC}, and point-wise uniqueness comes from the conjunctive closure in C^0_{FUNC} and the presuppositional Ans_{Dayal}-operator. This account also manages to keep the semantic type of questions low (i.e., single/double-*wh* questions and ∀-questions are uniformly of type $\langle st, t\rangle$), leaving space to tackle *wh*-constructions that are more complex (e.g., *wh*-triangles, multi-*wh* echo questions).

However, this account faces many problems. Conceptually, the composition involves a few *ad hoc* or problematic assumptions. First, the index abstractions are isolated from the moved *wh*-phrases and quantifiers. This way of abstracting indices is especially concerning since here the structure involves multiple abstractions — isolating the λ-operators from the moved phrases make the binding relations ambiguous. Second, the C^0_{FUNC} is structure specific and is hard-wired with a complex semantics. It is unclear why a covert functional head should be interpreted as such and appear only in particular structures. Thus, Dayal is not fully satisfied with this approach and calls it the "crazy C^0 approach." Last, for ∀-questions in specific, it is implausible to move a non-interrogative phrase to the specifier of an interrogative CP (Heim 2012).

In addition to the above conceptual problems, this account also makes a couple of problematic empirical predictions. (Note that these problems are independent from assuming functionality.) First of all, composing pair-list ∀-questions and multi-*wh* questions based on the very same LF, this account predicts that the two types of pair-list questions are semantically equivalent. However, as argued in section 2.1, the two questions differ with respect to domain exhaustivity. As seen in (10), repeated below, the multi-*wh* question, but not the ∀-question, can be felicitously used in a context where domain exhaustivity is violated.

(36) (Context: **100** *candidates are competing for* **three** *job openings.*)

a. "Guess which candidate will get which job."

b. # "Guess which job will every candidate get."

Second, this account does not extend to choice readings of ∃-questions. To avoid over-generating pair-list readings for ∃-questions (recall the limited distribution of pair-list from Sect. 3), Dayal stipulates that the quantification domain of a non-interrogative quantifier can only be obtained by extracting the unique minimal witness set of this quantifier. Table 1 illustrates the minimal witness sets of the three basic generalized quantifiers in a discourse domain with three boys *abc*. Observe that only the universal quantifiers have a non-empty unique witness set, which is simply their smallest live-on set. In contrast, existential indefinites have multiple minimal witness sets. Negative quantifiers (and other decreasing quantifiers) have a unique minimal witness set, which is however the empty set. With this stipulation, the LF (33) assumed for composing pair-list questions is unavailable for questions with a non-universal quantifier. Although this stipulation avoids over-generating pair-list readings in questions with a non-universal quantifier, it is pretty *ad hoc* and leaves choice readings of ∃-questions unexplained.

(37) **Live-on sets and witness sets** (Barwise and Cooper 1981)
For any π of type $\langle et, t\rangle$:

a. π **lives on** a set B if and only if $\pi(C) \Leftrightarrow \pi(C \cap B)$ for any set C;

b. If π lives on B, then A is a **witness set** of π if and only if $A \subseteq B$ and $\pi(A)$.

Generalized quantifier π	Minimal witness set(s) of π
every/each boy	$\{a, b, c\}$
one of the boys	$\{a\}$, $\{b\}$, $\{c\}$
no boy	$\varnothing$

Table 1: Illustration of minimal witness sets (with three relevant boys *abc*)

Third, as pointed out by Lahiri (2002), defining a pair-list question as a set of conjunctive propositions, this account has difficulties in accounting for the QV effects in embeddings of pair-list questions. For example, the question-embedding sentence (38) implies a QV inference, which can be paraphrased as if the matrix quantificational adverbial *mostly* quantifies over a set of atomic propositions. However, these atomic propositions cannot be retrieved from the question denotation assumed in (31): from a conjunctive proposition, we cannot extract out its propositional conjuncts semantically.

(38) Jill mostly knows [$_{\text{PAIR-LIST}}$ { which movie every boy watched / which boy watched which movie }].
⇝ 'For most true propositions *p* of the form ⌜*boy-x watched movie-y*⌝, Jill knows *p*.'

To account for the QV effects, in an on-going work, Dayal (2016) proposes to get rid of the ⋂-closure in C^0_{FUNC} and analyze the root denotation of a pair-list question as a family of sets of propositions. The revised account manages to keep the atomic propositions alive, but it sacrifices the advantage of keeping the semantic type of questions low.

4.2. Family-of-questions approaches

Family-of-questions approaches regard a pair-list question as denoting a set/family of sub-questions (Hagstrom 1998; Preuss 2001; Fox 2012a,b; Nicolae 2013; Kotek 2014; Xiang 2016: chapter 5; Dayal 2016; among others). As exemplified in (39), if a simplex *wh*-question denotes a set of propositions, a family of questions denotes a set of sets of propositions.[10]

(39) (Context: *There are two relevant boys $b_1 b_2$ and two relevant movies $m_1 m_2$.*)
Which movie did every boy watch?/ Which boy watched which movie?

$$[\![\mathrm{Q}_\forall]\!] = [\![\mathrm{Q}_{\text{multi-}wh}]\!] = \{[\![\textit{which movie did x watch?}]\!] \mid x \in \mathsf{B}_@\}$$
$$= \{\{\lambda w.\mathsf{W}_w(x,y) \mid y \in \mathsf{M}_@\} \mid x \in \mathsf{B}_@\}$$
$$= \left\{ \begin{array}{l} \{\lambda w.\mathsf{W}_w(b_1,m_1), \lambda w.\mathsf{W}_w(b_1,m_2)\} \\ \{\lambda w.\mathsf{W}_w(b_2,m_1), \lambda w.\mathsf{W}_w(b_2,m_2)\} \end{array} \right\}$$

The non-flat semantics assumed in (39) makes it easy to account for the QV effects in embeddings of pair-list questions. As in (40), the QV inference can be defined as if the matrix adverbial *mostly* quantifies over a set of sub-questions.

(40) Jill mostly knows $[_{\text{PAIR-LIST}} \left\{ \begin{array}{l} \text{which movie every boy watched} \\ \text{which boy watched which movie} \end{array} \right\}]$.
⇝ 'For most Q s.t. Q is a question of the form ⌜*which movie did boy-x watch?*⌝, Jill knows Q.'

Fox (2012a,b) composes the two pair-list questions via different LFs that yield the very same root denotation. The LF of a pair-list multi-*wh* question is illustrated in (41). As *wh*-phrases are defined as existential indefinites (viz., [\[*which boy*]\] = [\[*some boy*]\]), this LF is read as 'the set of Q such that for some boy x, Q is identical to [\[*which movie did x watch?*]\].' This composition follows the Government and Binding style of Karttunen Semantics (Heim 1995) except that it treats the identity (Id-)operator type-flexible and allows this operator to be iterated.

(41) Which boy watch which movie? (Pair-list reading)

$[_{\mathrm{CP}_2}\ \lambda Q_{\langle st,t\rangle}$ [wh-boy$_@$ λx_e $[_{\mathrm{C}'_2}$[Id Q] $[_{\mathrm{CP}_1}$ λp_{st} [wh-movie$_@$ λy_e $[_{\mathrm{C}'_1}$[Id p] $[_{\mathrm{IP}}$ x watch y]]]]]]]

a. $[\![\mathrm{Id}]\!] = \lambda\alpha_\tau\lambda\beta_\tau.\alpha = \beta$ (τ stands for an arbitrary type)

[10] The analyses of Groenendijk and Stokhof (1984) and Chierchia (1993) are also family-of-questions approaches. They define a QiQ-question as a family of sub-questions that quantify over a minimal witness set (mws) of the involved generalized quantifier $\mathcal{P}$, as schematized in (i).

(i) $[\![\textit{which movie did } \mathcal{P}_{boy} \textit{ watch?}]\!]_{\mathrm{QiQ}} = \{[\![\textit{which member of A watched which movie?}]\!] \mid \textsc{mws}(\mathcal{P}, A)\}$

However, the predictions made by these two accounts are quite different from the predictions of the non-flat semantics in (39). For example, Chierchia (1993) defines the sub-question as a set of propositions of the form ⌜*boy-x watched movie* $\mathbf{f}(x)$⌝, as in (ii). The denotations of the related ∀/∃-questions are thus illustrated as in (iii). Chierchia further assumes that answering a family of sub-questions means answering <u>one</u> of the sub-questions (in contrast to Fox's assumption that answering a family of sub-questions means answering <u>all</u> of the sub-questions.). Since the existential quantifier *one of the boys* has multiple minimal witness sets, the ∃-question has a choice flavor. While this account naturally extends to ∃-questions, it cannot explain the effects of ∀-questions such as domain exhaustivity, point-wise uniqueness, and QV.

(ii) $[\![Q_{\mathcal{P}}]\!] = \{\{\lambda w.\mathsf{W}_w(x,\mathbf{f}(x)) \mid x \in A, \mathbf{f} \in [A \to \mathsf{B}_@]\} \mid \textsc{mws}(\mathcal{P}, A)\}$

(iii) (Context: *There are two relevant boys $b_1 b_2$ and two relevant movies $m_1 m_2$.*)

a. $[\![Q_\forall]\!] = \left\{\left\{ \begin{array}{l} \lambda w.\mathsf{W}_w(b_1,m_1), \lambda w.\mathsf{W}_w(b_2,m_2), \\ \lambda w.\mathsf{W}_w(b_1,m_2), \lambda w.\mathsf{W}_w(b_2,m_2) \end{array} \right\}\right\}$

b. $[\![Q_\exists]\!] = \left\{ \begin{array}{l} \{\lambda w.\mathsf{W}_w(b_1,m_1), \lambda w.\mathsf{W}_w(b_1,m_2)\}, \\ \{\lambda w.\mathsf{W}_w(b_2,m_1), \lambda w.\mathsf{W}_w(b_2,m_2)\} \end{array} \right\}$

b. $[\![\text{IP}]\!] = \lambda w.\mathsf{W}_w(x, y)$

c. $[\![\text{C}_1']\!] = [\![\text{ID}]\!](p)([\![\text{IP}]\!])$
$\quad = p = \lambda w.\mathsf{W}_w(x, y)$

d. $[\![\text{CP}_1]\!] = \lambda p.\exists y[\mathsf{M}_@(y) \wedge p = \lambda w.\mathsf{W}_w(x, y)]$
$\quad = \{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\}$

e. $[\![\text{C}_2']\!] = [\![\text{ID}]\!](Q)([\![\text{CP}_1]\!])$
$\quad = Q = \{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\}$

f. $[\![\text{CP}_2]\!] = \lambda Q.\exists x[\mathsf{B}_@(x) \wedge Q = \{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\}]$
$\quad = \{\{\lambda w.\mathsf{W}_w(x, y) \mid y \in \mathsf{M}_@\} \mid x \in \mathsf{B}_@\}$

The LF of the corresponding pair-list ∀-question is as in (42), read as 'the unique minimal set K such that for every boy x: $[\![$*which movie did x watch?*$]\!]$ is a member of K.' The most important operations involved in forming this LF are (i) **quantifying-into predication** and (ii) **minimization** (*à la* Pafel 1999; Preuss 2001). For operation (i), the ∀-subject undergoes quantifier raising and quantifies into a predication condition yielded by applying a null predicative variable K to an open *wh*-question. This operation yields a universal predication condition, read as 'for every boy x: $[\![$*which movie did x watch?*$]\!]$ is a member of K'. For operation (ii), the MIN-operator binds the K variable across the subject-quantifier *every boy*, returning the unique minimal K set that satisfies the universal predication condition. This minimal set is simply the set consisting of all the questions of the form ⌜*which movie did boy-x watch?*⌝.

(42) Which movie did every boy watch? (Pair-list reading)

$[_{\text{CP}_2}$ MIN $\lambda \mathsf{K}_{\langle st,t\rangle}$ [every-boy$_@$ λx_e [K $[_{\text{CP}_1}$ λp_{st} [wh-movie$_@$ λy_e [[ID p] $[_{\text{IP}}$ x watch y]]]]]]]

a. $[\![\text{CP}_1]\!] = \{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\}$ (Composition is the same as in (41a-d))

b. $[\![\text{MIN}]\!] = \lambda \alpha_{\langle \sigma t,t\rangle} : \exists \mathsf{K}_{\langle \sigma,t\rangle}[\mathsf{K} \in \alpha \wedge \forall \mathsf{K}' \in \alpha[\mathsf{K} \subseteq \mathsf{K}']].\iota \mathsf{K}_{\langle \sigma,t\rangle}[\mathsf{K} \in \alpha \wedge \forall \mathsf{K}' \in \alpha[\mathsf{K} \subseteq \mathsf{K}']]$
(For a set of sets α, $[\![\text{MIN}]\!](\alpha)$ is the unique minimal set in α which is a subset of every set in α, defined only if this minimal set exists.) (Pafel 1999)

c. $[\![\text{CP}_2]\!] = [\![\text{MIN}]\!](\lambda \mathsf{K}.[\![\textit{every boy}_@]\!](\lambda x.\mathsf{K}(\{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\})))$
$\quad = [\![\text{MIN}]\!](\lambda \mathsf{K}.\forall x[\mathsf{B}_@(x) \rightarrow \mathsf{K}(\{\lambda w.\mathsf{W}_w(x, y) \mid \mathsf{M}_@(y)\})])$
$\quad = \{\{\lambda w.\mathsf{W}_w(x, y) \mid y \in \mathsf{M}_@\} \mid x \in \mathsf{B}_@\}$

As for answerhood, Fox (2012a,b) assumes that answering a family of sub-questions means answering each of the contained sub-questions. In other words, the answerhood operation is applied point-wise. As recursively defined in (43), when applied to a set of sub-questions, the point-wise answerhood-operator imposes ANS_{Dayal} to each sub-question and returns the conjunction of the strongest true propositional answer to each sub-question. Since the fronted *wh*-phrase is singular-marked, the point-wise triggered presupposition that every sub-question has a strongest true answer yields domain exhaustivity and point-wise uniqueness.

(43) Point-wise answerhood-operator (Fox 2012a)

$$\text{ANS}_{\text{PW}} = \lambda w \lambda Q. \begin{cases} \text{ANS}_{Dayal}(w)(Q) & \text{if } Q \text{ is of type } \langle st, t\rangle \\ \bigcap\{\text{ANS}_{\text{PW}}(w)(\alpha) \mid \alpha \in Q\} & \text{otherwise} \end{cases}$$

Fox's account has two advantages over the account of Dayal (1996, 2017). First, as discussed in (40), the non-flat semantics of pair-list questions can easily account for the QV effects in embeddings. Second, the composition is quite neat; it does not use any *ad hoc* composition rules or type-shifting rules or employ any complex operators. In composing the multi-*wh* question, the same as assumed in

Karttunen Semantics, the *wh*-phrases function as existential indefinites and quantify into an identity condition. In composing the ∀-question, the subject-quantifier standardly combines with a one-place predicate (of type $\langle e, t \rangle$) denoted by its sister node; hence, there is no need to stipulate a type-shifting operation to extract the quantification domain of the quantifier.

However, the account of Fox is subject to the same empirical problems as the account of Dayal (1996, 2017). First, treating pair-list ∀-questions semantically equivalent to their multi-*wh* counterparts, Fox also cannot explain the contrast with respect to domain exhaustivity. Second, this account does not extend to ∃-questions either. In composing questions with quantifiers, Fox uses the MIN-operator to obtain the unique minimal K set that satisfies the quantificational predication condition, which is however unavailable if the predication is existentially quantified. For instance, for the ∃-question (44a), in a discourse with two relevant boys b_1 and b_2, the smallest K sets satisfying the existential quantification condition (44b) are the two sets in (44c), neither of which is a subset of the other.

(44) a. Which movie did one of the boys watch?
b. $\exists x[\mathsf{B}_@(x) \wedge [\![\textit{which movie did x watch?}]\!] \in \mathsf{K}]$
c. $\{[\![\textit{which movie did } b_1 \textit{ watch?}]\!]\}, \{[\![\textit{which movie did } b_2 \textit{ watch?}]\!]\}$

5. Formal theory: A hybrid categorial approach

My general treatment of question composition follows the hybrid categorial approach developed in Xiang 2016, 2020. This approach follows traditional categorial approaches in assuming that questions denote functions but overcomes their technical problems in composition. Compared with proposition-based frameworks such as Hamblin-Karttunen Semantics, this framework allows to derive QV effects in embeddings of pair-list questions without having to assume a non-flat semantics (Sect. 7). Note that, however, assumptions made in Sect. 6 on the composition of the question nucleus are independent from this framework.

The hybrid categorial approach has three main ingredients. First, matrix and embedded questions uniformly denote functions from short answers to corresponding propositional answers, called "topical properties". For example, the question in (45) denotes a function that maps each atomic boy x to the proposition that x came. In consequence, short answers are extractable from question denotations as meanings in the property domain. This assumption is basic in any categorial approach to questions. It will be crucial for explaining the QV effects in embeddings of pair-list questions.

(45) a. $[\![\textit{which boy came?}]\!] = \lambda x_e : \mathsf{B}_@(x) .\lambda w[\mathsf{C}_w(x)]$
b. $[\![\textit{which boy came?}]\!]([\![\textit{John}]\!]) = \mathsf{B}_@(j) .\lambda w[\mathsf{C}_w(j)]$

Second, *wh*-phrases are existential quantifiers ranging over polymorphic sets. In questions with extensional readings, the quantification domain of a *wh*-phrase of the form $\ulcorner wh\text{-}\mathrm{A}_w \urcorner$ consists of not only elements in the extension of the *wh*-complement $[\![\mathrm{A}]\!]^w$ but also Skolem functions to $[\![\mathrm{A}]\!]^w$, as defined in (46b). The semantics of *wh*-phrases in questions with an intensional reading is defined analogously, as schematized in (46c).

(46) **The semantics of a *wh*-phrase** (Modified from Xiang 2020)
a. For any set A, $\mathrm{Ran}(\mathbf{f}) \subseteq A$ if and only if $\forall x \in \textsc{Dom}(\mathbf{f})[\mathbf{f}(x) \in A]$.

b. For extensional readings

$$[\![wh\text{-}A_w]\!] = \lambda P.\exists\alpha \in \bigcup \left\{ \begin{array}{l} [\![A]\!]^w, \\ \{\mathbf{f} \mid \text{Ran}(\mathbf{f}) \subseteq [\![A]\!]^w\} \end{array} \right\} [P(\alpha)]$$

c. For intensional readings

$$[\![wh\text{-}\lambda w.A_w]\!] = \lambda P.\exists\alpha \in \bigcup \left\{ \begin{array}{l} \{R \mid \forall w[R(w) \in [\![A]\!]^w]\}, \\ \{\mathbf{f} \mid \forall w[\text{Ran}(\mathbf{f}(w)) \subseteq [\![A]\!]^w]\} \end{array} \right\} [P(\alpha)]$$

The above definitions treat *wh*-expressions as existential indefinites. In the composition of a *wh*-question, however, fronted *wh*-phrases are type-shifted into type-flexible function domain restrictors via the application of a BeDom-operator. For any existential quantifier π, Be(π) is the set that π ranges over (Partee 1986), and BeDom(π) is a function domain restrictor which combines with a function θ and returns the function that is similar to θ but is undefined for items not in Be(π).

(47) **The BeDom-operator**

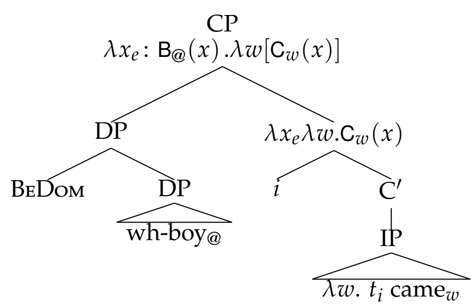
For any π of type $\langle \sigma t, t\rangle$ where σ is an arbitrary type, we have:

a. $\text{Be}(\pi) = \lambda x.\mathcal{P}(\lambda y.y = x)$

b. $$\text{BeDom}(\pi) = \lambda\theta_\tau.\iota P_\tau \left[\begin{array}{r} [\text{Dom}(P) = \text{Dom}(\theta) \cap \text{Be}(\pi)] \\ \wedge\ \forall\alpha \in \text{Dom}(P)[P(\alpha) = \theta(\alpha)] \end{array} \right]$$

For example, in the LF (48), 'BeDom (wh-boy$_@$)' combines with the 'came'-property defined for all entities and returns the 'came'-property defined only for entities that are atomic boys. The same as discussed in footnote 7, LF representations are translated into Ty2. World variables of predicative expressions within the nucleus are abstracted at the edge of IP. The extension of the *wh*-restrictor is evaluated relative to the actual world @.

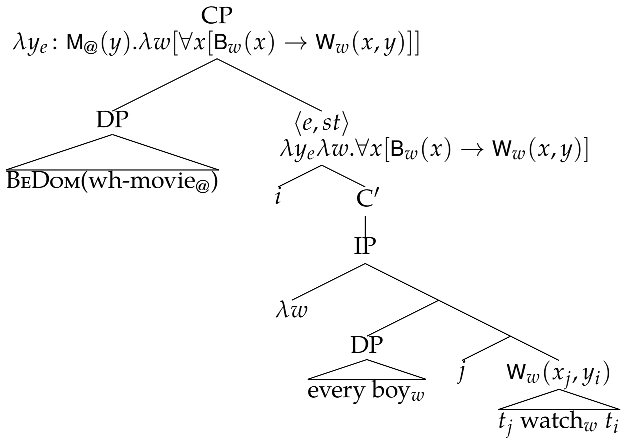
(48) Which boy came?



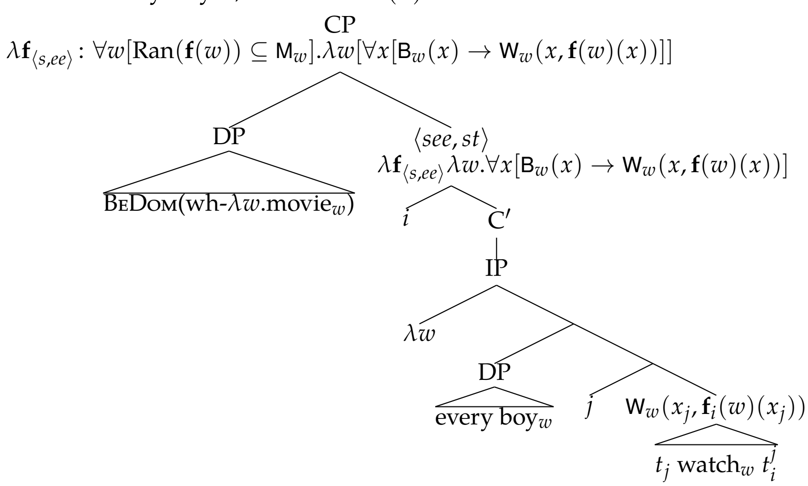
Crucially, BeDom(π) is type-flexible — it can combine with any function of a $\langle \sigma, ...\rangle$ type where σ is the type of an element in Be(π). Type-flexibility makes it possible to compose a question regardless of whether the function denoted by the question nucleus is defined for individuals or functions, and regardless of how many *wh*-phrases there are in this question. Take the single-*wh* question (49) for example. This question has an individual reading if the fronted *wh*-phrase binds an individual trace, as in (49a), and a functional reading if it binds an (intensional) functional trace, as in (49b). As for multi-*wh* questions, the tree diagram in (50) illustrates the derivation of single-pair readings. 'BeDom(wh-movie$_@$)' applies to a one-place property of type $\langle e, st\rangle$ defined for any entities and returns a similar property defined only for atomic movies. Likewise, 'BeDom(wh-boy$_@$)' applies to a two-place property of type $\langle e, \langle e, st\rangle\rangle$ defined for any entities and returns a similar property defined only for atomic boys.

(49) Which movie did every boy watch?

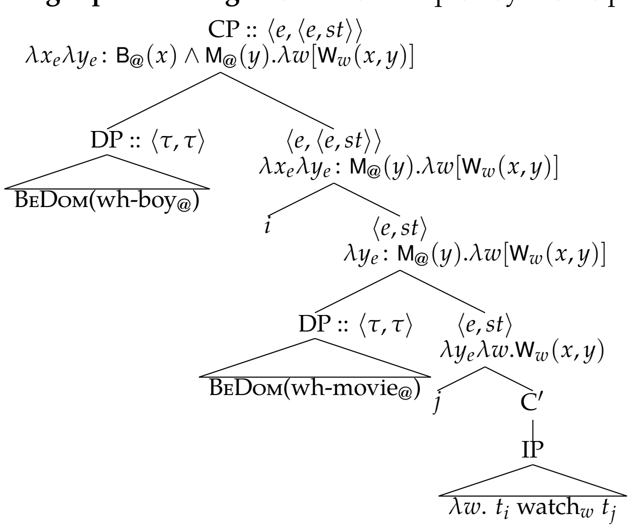
a. **Individual reading**: 'Which movie y is such that every boy watched y?'



b. **(Intensional) functional reading**: 'Which Skolem function **f** to atomic movies is such that for every boy x, x watched $\mathbf{f}(x)$?'



(50) Which boy watched which movie?
Single-pair reading: 'For which unique boy-movie pair $\langle x, t \rangle$ is such that x watched y?'



Last, complete true answers of questions are obtained by applying the answerhood-operators in (51). Compared with the Ans_{Dayal}-operator (34), the major difference is that the Hamblin set Q is replaced with a topical property $\boldsymbol{P}$, which can supply both propositional answers and short answers.[11] These answerhood operators account for uniqueness effects in the same way as Ans_{Dayal}.

(51) **Answerhood-operators**

a. For the complete true short answer

$$\text{Ans}^S(w)(\boldsymbol{P}) = \exists\alpha \in \text{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\alpha) \wedge \forall\beta \in \text{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\beta) \rightarrow \boldsymbol{P}(\alpha) \subseteq \boldsymbol{P}(\beta)]].$$
$$\iota\alpha \in \text{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\alpha) \wedge \forall\beta \in \text{Dom}(\boldsymbol{P})[w \in \boldsymbol{P}(\beta) \rightarrow \boldsymbol{P}(\alpha) \subseteq \boldsymbol{P}(\beta)]]$$

b. For the complete true propositional answer

$$\text{Ans}(w)(\boldsymbol{P}) = \boldsymbol{P}(\text{Ans}^S(w)(\boldsymbol{P}))$$

6. Proposal

In line with functionality-based approaches, I analyze pair-list readings of multi-*wh* questions and QiQ-readings of questions with quantifiers as extensional functional readings. For both types of questions, I assume that the composition involves a quantificational condition with respect to an open sentence of the form $\ulcorner x \text{ P } \mathbf{f}(x) \urcorner$ which expresses a functional dependency relation between the two arguments of a two-place predicate 'P'. In particular, in a pair-list multi-*wh* question, the composition involves existential quantification of the subject-*wh* into an identity condition (*à la* Karttunen Semantics); while in a QiQ-question, the composition involves that the subject-quantifier quantifies into a predication condition (*à la* Fox 2012b). A general schema of composition is as follows, repeated from (6):

(52) **A general schema of composing complex questions**

a. Which boy watched which movie? (Pair-list reading)

... [which-movie$_j$ ... which-boy$_i$ [$_{\text{IDENT}}$ ... [$_{\text{VP}}$ t_i watched t_j^i]]]

b. Which movie did Det-boy(s) watch? (QiQ-reading)

... [which-movie$_j$ ... Det-boy(s)$_i$ [$_{\text{PRED}}$ ... [$_{\text{VP}}$ t_i watched t_j^i]]]

The subtle distinctions between these two operations of quantifying into functional dependencies lead to a sharp contrast between multi-*wh* and ∀- questions with respect to domain exhaustivity. In addition to ∀-questions, the composition schema for QiQ-readings also automatically explains why questions with existential indefinites have choice readings as well as why those with negative questions do not have QiQ-readings. What's more, with known contrasts among non-interrogative quantifiers with respect to lexical distributivity and productivity of scoping, this analysis can also explain why counting quantifiers do not license QiQ-readings.

In what follows, I will provide the root denotation of each type of complex questions up front (Sect. 6.1) and then show how to derive these root denotations compositionally (Sect. 6.2 and 6.3).

[11] Following Fox (2013), Xiang (2016, 2020) assumes a weaker definition of complete answers: a true answer to a question is complete as long as it is not asymmetrically entailed by any true answers to this question. This answerhood is assumed to account for mention-some readings of questions and free relatives. Since mention-some is not the focus of this paper, for easier comparisons with competing theories in composing complex questions, here I follow Dayal (1996, 2017) and define the complete true answer as the unique strongest true answer.

6.1. Question denotations

I propose that pair-list readings and QɪQ-readings of complex *wh*-questions are extensional functional readings. When having a QɪQ/pair-list reading, a question denotes a topical property of type $\langle ee, st\rangle$ that maps a Skolem function to the conjunctive proposition that expresses the graph of this Skolem function. Formal illustrations of the denoted topical properties are given in (53-54) in tandem. In specific, the (a)-denotations are represented in a way isomorphic to the structures of composition (for details of composition, see the rest of Sect. 6), and the (b)-denotations are semantically equivalent to their (a)-counterparts but are represented in a way more convenient for comparison.

(53) $[\![\textit{which boy watched which movie?}]\!]_{\text{pair-list}}$
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ . \bigcap\{p \mid \exists\text{-}\mathsf{B}_@(\lambda x.p = \lambda w.\mathsf{W}_w(x, \mathbf{f}(x)))\}$ (a)
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ . \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}]$ (b)

(54) $[\![\textit{which movie did Det-boy(s) watch?}]\!]_{\text{QiQ}}$
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ . \bigcap \text{E-}\textsc{min}(\{\mathsf{K} \mid \mathsf{Det\text{-}B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))\})$ (a)
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ \wedge \mathsf{Det\text{-}B}_@(\text{Dom}(\mathbf{f})).$
$\bigcap \text{E-}\textsc{min}(\{\mathsf{K} \mid \mathsf{Det\text{-}B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))\})$ (b)

A crucial contrast between (53b) and (54b) is that the former restricts only the range of the input Skolem functions, while the latter restricts also the domain. More specifically, in (53), the topical property of the multi-*wh* question maps any Skolem function that maps entities to atomic movies to the graph of this function. In contrast, in (54), the topical property yielded by the corresponding QiQ-question is defined more restrictively only for Skolem functions that map Det-boy(s) to atomic movies, and this topical property maps each such Skolem function to the conjunction of a proposition set that quantifies over exactly Det-boy(s). The additional domain restriction in (54b), namely $\mathsf{Det\text{-}B}_@(\text{Dom}(\mathbf{f}))$, is a definedness condition of the value description in (54a): the quantificational predication condition $\mathsf{Det\text{-}B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))$, read as 'for Det-boy(s) x, the proposition $\ulcorner x$ *watched* $\mathbf{f}(x)\urcorner$ is a member of K', is defined only if the Skolem function $\mathbf{f}$ is defined for Det-boy(s).

For a more concrete illustration of the QɪQ-denotation, consider the related ∀-question. If the 'Det' in (54) is *every*/*each*, the defined topical property is as follows:

(55) $[\![\textit{which movie did every/each boy watch?}]\!]$
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \underbrace{\text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@}_{\text{from } wh\text{-obj}} . \underbrace{\bigcap \text{E-}\textsc{min}(\{\mathsf{K} \mid \forall\text{-}\mathsf{B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))\})}_{\text{(i) from nucleus}}$ (a)
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \underbrace{\text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@}_{\text{from } wh\text{-obj}} . \underbrace{[\forall\text{-}\mathsf{B}_@(\text{Dom}(\mathbf{f})). \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}]}_{\text{(ii) equivalent to (i)}}$ (a')
$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \underbrace{\text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@}_{\text{from } wh\text{-obj}} \wedge \underbrace{\forall\text{-}\mathsf{B}_@(\text{Dom}(\mathbf{f}))}_{\text{from (ii)}} . \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}$ (b)

In (55a), the input can be any function $\mathbf{f}$ from entities to atomic movies, and for each such input $\mathbf{f}$, the output is the conjunction of the set that consists of exactly <u>all</u> propositions of the form $\ulcorner$*boy-x watched* $\mathbf{f}(x)\urcorner$. Crucially, as represented explicitly in (55a'), this output inference is partial — the universal predication condition over the open sentence 'boy-x watched $\mathbf{f}(x)$' is defined only if $\mathbf{f}$ is defined for every boy in the discourse domain, which therefore yields domain exhaustivity. Finally, as in (55b), moving this definedness condition to the domain condition of the topical property yields that the input Skolem functions pair <u>every</u> boy with an atomic movie. In short, the topical property of the

∀-question is the same as that of the corresponding multi-*wh* question, except that it presupposes domain exhaustivity.

At this point, it should be clear why I pursue a functionality-based approach instead of a family-of-questions approach: the domain exhaustivity effect in a ∀-question comes from a definedness condition of applying quantification into functional dependency. In family-of-question approaches, however, domain exhaustivity is attributed to an operation outside the question nucleus (e.g., the point-wise answerhood-operator as in the analysis of Fox 2012a,b), which clearly cannot capture the semantic contrast between ∀-questions and multi-*wh* questions in terms of their structural differences.

6.2. Composing pair-list multi-*wh* questions

The tree diagram in Figure 2 illustrates the derivation of the root denotation of a pair-list multi-*wh* question. As marked in the tree diagram, this composition precedes in four steps. **First, deriving functional dependency.** Within IP, the argument variable of the complex functional trace of the object-*wh* is co-indexed with the trace of the subject-*wh*, yielding an open proposition that expresses a functional dependency relation between the subject and object arguments of *watch*. **Second, quantifying-into an identity condition.** An identity (Id-)operator yields an identity relation between a covert variable p and the open sentence denoted by IP. At node 1, the subject-*wh*, interpreted as an existential quantifier, binds the argument variable in IP across the Id-operator, yielding an existential identity condition with respect to a sentence expressing functional dependency. **Third, creating a function graph.** Abstracting the variable p cross the existential identity condition yields the set of propositions of the form $\ulcorner$*boy-x watched* $\mathbf{f}(x)\urcorner$. Conjoining this set of propositions by a $\bigcap$-closure yields the graph of the Skolem function $\mathbf{f}$. This $\bigcap$-closure can be considered as a *function graph creator* (Fgc) in the sense of Dayal (2017). **Last, creating a topical property**. Abstracting the index of the functional variable yields a function (of type $\langle ee, st\rangle$) that maps each Skolem function to a proposition that describes the graph of this Skolem function. Further, the fronted DP 'BeDom(wh-movie$_@$)' restricts the domain of this function and yields a similar function only defined for Skolem functions that range over atomic movies. The yielded function is the topical property of the multi-*wh* question.

(56) Steps 1 & 2: Quantifying-into an identity condition of a functional dependency

a. $[\![\text{IP}]\!] = \lambda w.\mathsf{W}_w(x_j, \mathbf{f}_i(x_j))$

b. $[\![\text{Id}]\!] = \lambda\alpha_\tau\lambda\beta_\tau.\alpha = \beta$

c. $[\![\text{C}']\!] = [\![\text{Id}(p)]\!]([\![\text{IP}]\!])$
$= p = \lambda w.\mathsf{W}_w(x_j, \mathbf{f}_i(x_j))$

d. $[\![\textit{which boy}_@]\!] = \lambda P_{\langle e,t\rangle}.\exists x[\mathsf{B}_@(x) \wedge P(x)]$

e. $[\![1]\!] = [\![\textit{which boy}_@]\!]([\![\text{C}']\!])$
$= \exists x[\mathsf{B}_@(x) \wedge p = \lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x))]$

(57) Step 3: Creating a function graph

a. $[\![\text{CP1}]\!] = \lambda p.\exists x[\mathsf{B}_@(x) \wedge p = \lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x))]$
$= \{\lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x)) \mid \mathsf{B}_@(x)\}$

b. $[\![2]\!] = \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x)) \mid \mathsf{B}_@(x)\}$

(58) Step 4: Creating a topical property
$[\![\text{CP2}]\!] = \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@. \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}$

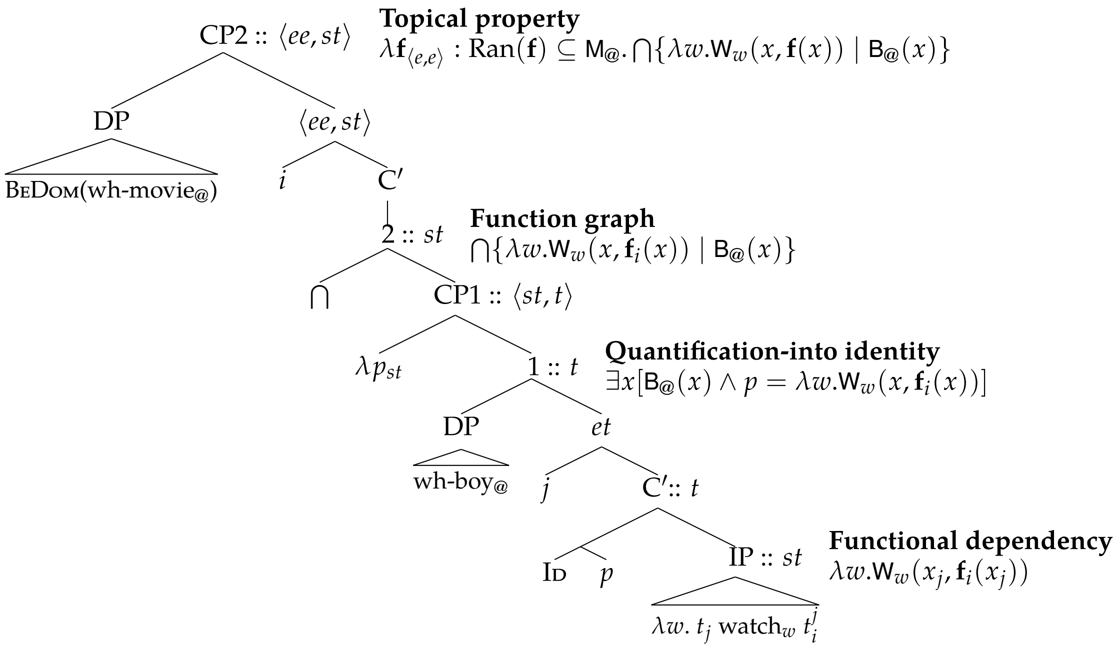


Figure 2: Composition of the pair-list multi-*wh* question *which boy watched which movie?*

It is worthy noting that, in contrast to basic functional questions, pair-list multi-*wh* questions do not admit fragment functional answers like (59a). Instead, multi-*wh* questions are only congruent with fragment answers that are lists of pairs of type $\langle se, se\rangle$ as in (59b) (Kang 2012; Sharvit and Kang 2017).[12] From the perspective of functionality-based approaches, as Chierchia (1993) argues, this gap shows that pair-list readings can be treated as special functional readings, but functional readings cannot be treated as special pair-list readings because the distribution of functional readings is more restrictive.

(59) Which boy watched which movie?

a. # His favorite superhero movie.

b. Andy, *Ironman*, Billy, *Spiderman*, Clark, *Hulk*.

6.3. Composing QiQ-questions

The root denotation of the QiQ-question in (54) is uniformly composed based on the LF schema in Figure 3. In particular, as for the denotation in (54b), the condition on the range of the input Skolem functions (i.e., **f** maps to atomic movies) is supplied by the fronted *wh*-phrase. All the rest, including the condition on the domain of the input Skolem function (i.e., that **f** is defined for Det-boy(s)) and the output proposition, are from the question nucleus (i.e., the scope of the fronted *wh*-phrase). Observe that the four general steps in this composition are in tandem with those in the composition of a pair-list multi-*wh* question. The following subsections will show how this composition schema derives each type of QiQ-readings.

[12] Sharvit and Kang (2017) provide an explanation to why pair-list questions do not admit intensional functional answers. However the syntax of multi-*wh* questions assumed by Sharvit and Kang is quite different from mine. I leave this issue open.

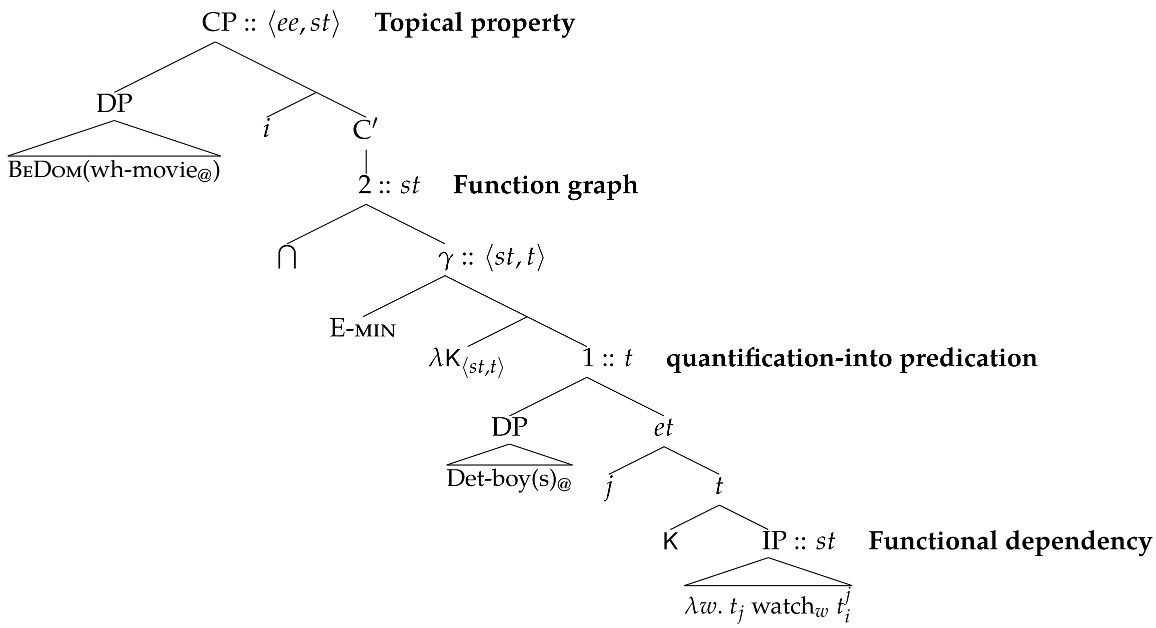


Figure 3: Composition of the QiQ-question *which movie did Det-boy(s) watch?*

(54) $[\![\textit{which movie did Det-boy(s) watch?}]\!]_{\text{QiQ}}$

$$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \underbrace{\text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@}_{\text{from } wh\text{-object}} . \underbrace{\bigcap \text{E-}\textsc{min}(\{\mathsf{K} \mid \mathsf{Det\text{-}B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))\})}_{\text{(i) from nucleus}} \quad \text{(a)}$$

$$\Leftrightarrow \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ \wedge \underbrace{\mathsf{Det\text{-}B}_@(\text{Dom}(\mathbf{f}))}_{\text{definedness cond of (i)}} . \bigcap \text{E-}\textsc{min}(\{\mathsf{K} \mid \mathsf{Det\text{-}B}_@(\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))))\}) \quad \text{(b)}$$

Recall that questions with quantifiers admit both functional readings and QiQ-readings. The following compares the derivations of QiQ-readings and basic functional readings.

On the one hand, the same as basic functional readings, pair-list readings of questions involve a functional dependency relation between the subject-quantifier and the object-*wh*. To derive this dependency, the fronted object-*wh* 'BeDom(*which movie*$_@$)' leaves a complex functional trace, whose argument index is bound by the subject-quantifier 'Det-*boy(s)*$_@$'.

On the other hand, different from the case of a basic functional reading but the same as in a pair-list multi-*wh* question, here the functional variable **f** in the complex functional trace is extensional (of type $\langle e, e\rangle$, not $\langle s, ee\rangle$). Moreover, here the nucleus involves two covert operations — **predication** and **minimization**. These operations are similar to what Fox (2012b) assumes for composing ∀-questions (see (42)), but they depart from Fox's particular implementation in two respects, yielding desirable consequences in accounting for the domain exhaustivity effects in ∀-questions and the choice readings of ∃-questions. First, in the presented analysis, the predication operation is applied to an open proposition $\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))$ (as opposed to an open question). This proposition expresses a functional dependency between the arguments of *watch*. The binding of the variables x and **f** contribute to the derivation of domain exhaustivity (Sect. 6.3.1). Second, the minimization operator E-min is semantically weaker than the min-operator that Fox adopts from Pafel (1999). As defined in (60) and illustrated in (61), the E-min-operator is lexically encoded with a choice function variable

$f_{\textsc{ch}}$ and does not presuppose uniqueness.[13] Replacing MIN with E-MIN makes the analysis extendable to ∃-questions (Sect. 6.3.2).

(60) $[\![\text{E-MIN}]\!] = \lambda\alpha_{\langle\sigma t,t\rangle}.f_{\textsc{ch}}(\{\mathsf{K}_{\langle\sigma,t\rangle} \mid \mathsf{K} \in \alpha \wedge \forall\mathsf{K}' \in \alpha[\mathsf{K}' \not\subset \mathsf{K}]\})$
(For a set of sets α: $[\![\text{E-MIN}]\!](\alpha)$ is a set K s.t. K is in α and no set in α is a proper subset of K. [$f_{\textsc{ch}}$ stands for a free choice function variable.])

(61) Let a and b be two distinct entities, $A = \{\varnothing, \{a\}, \{b\}\}$, and $B = \{\{a\}, \{b\}\}$. Then we have:
a. $[\![\text{MIN}]\!](A) = [\![\text{E-MIN}]\!](A) = \varnothing$;
b. $[\![\text{MIN}]\!](B)$ is undefined;
c. $[\![\text{E-MIN}]\!](B)$ has two possible values: $\{a\}$ and $\{b\}$.

6.3.1. Composing ∀-questions

This section presents the details of composing a pair-list ∀-question. The most important issues are to derive the pair-list reading and to account for the domain exhaustivity effect.

The LF is given in Figure 4. I divide the composition into four steps, in parallel to the composition of the corresponding pair-list multi-*wh* question (Sect. 6.2). **First, deriving functional dependency.** The IP denotes an open proposition expressing a functional dependency relation, composed in exactly the same way as the IP in the corresponding multi-*wh* question. **Second, quantifying-into a predication condition.** A null predicate K (of type $\langle\sigma, \sigma t\rangle$ where σ is an arbitrary type) combines with the open proposition denoted by IP, yielding a simple predication condition that this open proposition is a member of K. Next, the subject-quantifier *every/each-boy*$_{@}$ quantifies into this predication condition, yielding a universal predication condition as stated in (62b). Crucially, this universal predication condition is defined only if **f** is defined for every boy, which yields domain exhaustivity. **Third, creating a function graph.** Abstracting the predicative variable K returns the set of K sets that satisfy the universal predication condition yielded from Node 1. These are the sets that contain all the propositions of the form ⌜*boy-x watched* $\mathbf{f}(x)$⌝, as in (63a). At Node γ, applying the minimizer E-MIN returns one of the safistied minimal K sets. Among those satisfying the universal quantification predication, there is only one minimal set, namely, the set of the propositions of the form ⌜*boy-x watched* $\mathbf{f}(x)$⌝, as in (63b). At Node 2, this set of propositions is flattened by the application of a $\bigcap$-closure, returning a conjunctive proposition describing the graph of the function **f**, as in (63c). **Forth, creating a topical property.** The fronted 'BeDom(wh-movie$_{@}$)' binds the **f** variable and restricts the range of **f** to the set of atomic movies. The possible inputs of this topical property are therefore Skolem functions that map each boy to an atomic movie, and the outputs are conjunctive propositions describing the graph of this function.

(62) Step 1 & 2: Quantifying-into the predication condition of functional dependency
a. $[\![\text{IP}]\!] = \lambda w.\mathsf{W}_w(x_j, \mathbf{f}_i(x_j))$ (Equivalent to (56))
b. $[\![1]\!] = [\![\textit{every boy}_{@}]\!](\lambda x.\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x))))$
$= \forall x \in \mathsf{B}_{@}[\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}_i(x)))]$ (defined only if $\forall x \in \mathsf{B}_{@}[x \in \text{Dom}(\mathbf{f})]$)
(For every boy x, the proposition 'x watched $\mathbf{f}(x)$' is a member of K.)

(63) Step 3: Creating a function graph

[13] For readers who are familiar with Boolean Semantics, the E-MIN-operator is roughly the same as the collectivity raising operator in Winter 2001.

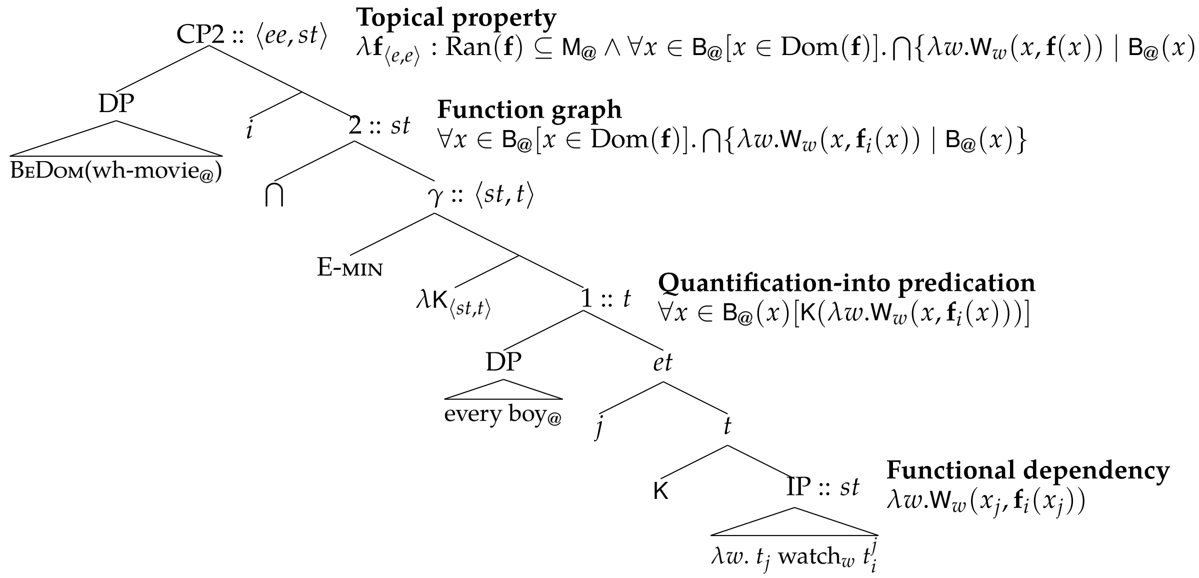


Figure 4: Composition of the ∀-question *which movie did every boy watch?*

a. $\lambda \mathsf{K}.[\![1]\!] = \lambda \mathsf{K}.\ \forall x \in \mathsf{B}_{@}[\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \in \mathsf{K}]$
$\quad = \lambda \mathsf{K}\colon \forall x \in \mathsf{B}_{@}[x \in \mathrm{Dom}(\mathbf{f})].\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_{@}(x)\} \subseteq \mathsf{K}$

b. $[\![\gamma]\!] = [\![\text{E-MIN}]\!](\lambda \mathsf{K}.[\![1]\!])$
$\quad = \forall x \in \mathsf{B}_{@}[x \in \mathrm{Dom}(\mathbf{f})].\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_{@}(x)\}$

c. $[\![2]\!] = \bigcap([\![\text{E-MIN}]\!](\lambda \mathsf{K}.[\![1]\!]))$
$\quad = \forall x \in \mathsf{B}_{@}[x \in \mathrm{Dom}(\mathbf{f})].\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_{@}(x)\}$

(64) Step 4: Creating a topical property
$[\![\mathrm{CP}]\!] = \lambda \mathbf{f}_{\langle e,e\rangle}\colon \mathrm{Ran}(\mathbf{f}) \subseteq \mathsf{M}_{@} \wedge \forall x \in \mathsf{B}_{@}[x \in \mathrm{Dom}(\mathbf{f})].\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_{@}(x)\}$

Step 2 of this composition — quantification-into predication — is especially important. First, it carries forward the advantage of Fox's analysis that the subject-quantifier standardly combines with a one-place predicate of type $\langle e, t\rangle$. In contrast to earlier accounts (e.g., Groenendijk and Stokhof 1984; Chierchia 1993; Dayal 1996, 2017), there is no need of assuming a type-shifting operation or making use of minimal witness sets. Moreover, since here the subject-quantifier also binds the argument variable of the functional trace, the quantifying-in operation yields a presupposition that the Skolem function **f** is defined for Det-boy(s). For example, for the ∀-question, the universal predication condition (62b) is defined only if **f** is defined for every boy. This presupposition projects over CP, yielding domain exhaustivity for the ∀-question.

The explanation of domain exhaustivity crucially relies on the presence of a universal quantifier — the domain exhaustivity effect comes from the universal predication condition, and especially, the binding relation between a <u>universal</u> quantifier and the argument of the functional *wh*-trace. Nicely, this analysis does not over-predict domain exhaustivity for a pair-list multi-*wh* question — in a multi-*wh* question, the argument variable of the functional trace of the object-*wh* is <u>existentially</u> bound by the subject-*wh*. For comparison, the family-of-questions approach of Fox (2012a,b) attributes domain exhaustivity to an operation outside the question nucleus, namely, the point-wise answerhood-operator. Since the choice of answerhood is independent from the root structure/meaning of a question, the family-of-questions approach cannot explain the contrast in domain exhaustivity between ∀-questions and multi-*wh* questions.

To sum up, the QɪQ-reading of a ∀-question is [+ᴅ-ᴇxʜ,+ᴘʟ,−ᴄʜ]. It is subject to domain exhaustivity because the universal predication condition (Node 1) is defined only if **f** is defined for every boy. It expects a pair-list answer because the yielded eligible minimal K set (Node γ) is a non-singleton set ranging over multiple boys. It does not have a choice flavor because there is only one minimal eligible K set.

6.3.2. Composing ∃-questions

The composition of a choice ∃-question is in analogy to that of the pair-list ∀-question. Note 1 creates an existential predication condition over the open proposition $\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))$, as in (65a). At Node γ, binding the K variable with the E-ᴍɪɴ-operator across the subject-indefinite *one of the boys* returns one of the minimal K sets that satisfy this existential predication condition. Crucially, different from the case of the ∀-question, here there are multiple eligible minimal K sets, each of which is a singleton set consisting of exactly one proposition of the form ⌜*boy-x watched* $\mathbf{f}(x)$⌝, as in (65b). ('$x = f_{\text{CH}}(\mathsf{B}_@)$' means that the boy x is chosen by a choice function variable f_{CH} encoded within the E-ᴍɪɴ-operator.) Each such minimal K set supplies a possible question denotation, which therefore gives rise to a choice flavor. The rest steps are the same as in the ∀-question.

(65) Which movie did one of the boys watch?

[$_{\text{CP}}$BᴇDᴏᴍ(wh-movie$_@$) $\lambda \mathbf{f}_{\langle e,e \rangle}$ [$_2$ $\bigcap$ [$_\gamma$ E-ᴍɪɴ $\lambda \mathsf{K}_{\langle st,t \rangle}$ [$_1$ one-boy$_@$ λx [K($\lambda w.x$-watch$_w$-$\mathbf{f}(x)$)]]]]]

a. $[\![1]\!] = \exists x \in \mathsf{B}_@[\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)))]$

b. $[\![\gamma]\!] = [\![\text{E-MIN}]\!](\lambda \mathsf{K}.[\![1]\!])$
$\quad = \{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))\}$, where $x = f_{\text{CH}}(\mathsf{B}_@)$

c. $[\![2]\!] = \bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))\}$
$\quad = \lambda w.\mathsf{W}_w(x, \mathbf{f}(x))$, where $x = f_{\text{CH}}(\mathsf{B}_@)$

d. $[\![\text{CP}]\!] = \lambda \mathbf{f}_{\langle e,e \rangle} : \text{Ran}(\mathbf{f}) \subseteq \mathsf{M}_@ .\lambda w[\mathsf{W}_w(x, \mathbf{f}(x))]$, where $x = f_{\text{CH}}(\mathsf{B}_@)$

In contrast to the case of a ∀-question, the yielded reading for an ∃-question is [−ᴅ-ᴇxʜ,−ᴘʟ,+ᴄʜ]. In specific, this reading is not subject to domain exhaustivity because the existential predication condition (65a) only requires **f** to be defined for at least one of the boys.[14] The possible answers to this question are single-pairs, not pair-lists, because the minimal K sets satisfying the existential predication condition are all singleton sets. This reading has a choice flavor because there can be multiple eligible minimal K sets satisfying the quantificational predication condition.

The above discussion is for the ∃1-quantifier *one of the boys*. The rest of this section extends this analysis to other existential indefinites of the form 'Num-NP' or 'Num-of-the-NP'. Recall from Sect. 3 that pair-list readings are not available in matrix ∃-question. For example, the ∃2-question in (66c) cannot be interpreted with distributivity in between quantification and uniqueness.

(66) I know that every student voted for a different candidate. Which candidate did ...

a. ... **every/each** student vote for? (∀/ᴇᴀᴄʜ $\gg \iota$)

b. ... **one** of the students vote for? (∃1 $\gg \iota$)

[14] In (65d-e), there is no need to write out the domain condition that **f** must be defined for at least one boy, because this condition is entailed by the definedness condition of the output proposition: for any chosen boy x, the proposition $\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))$ is defined only if **f** is defined for this x.

c. # ... **two** of the students vote for? ($\exists 2 \gg$ ~~EACH~~ $\gg \iota$)

To avoid over-generating pair-list readings, pioneering works such as Dayal 1996 and Fox 2012b derive pair-list readings in ways that would crash in questions with a non-universal quantifier. In Dayal's analysis, the derivation of pair-list crashes because existential quantifiers have multiple minimal witness sets. In Fox's analysis, the derivation crashes because we cannot find the unique minimal set among the sets that satisfy an existential predication condition over sub-questions. Obviously, this strategy comes with an expense of not being able to account for choice readings of ∃-questions.

I propose that the determiner of the numeral-modified indefinite *two of the boys* is not ∃2 but rather ∃; in other words, the cardinal numeral *two* is part of the restrictor of the determiner. With this assumption, the quantifier *two of the boys* ranges over the set of entities that are pluralities of two boys, and it denotes a set of sets that contain at least <u>one</u> of such plural entities.

(67) a. $\exists 2 =_{\text{def}} \lambda P_{\langle e,t\rangle} \lambda Q_{\langle e,t\rangle}.|P \cap Q| = 2$

b. $\exists =_{\text{def}} \lambda P_{\langle e,t\rangle} \lambda Q_{\langle e,t\rangle}.P \cap Q \neq \varnothing$

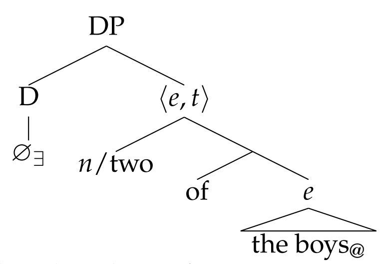
This assumption is supported by the contrast between (68a-b): unlike distributive universal quantifiers such as *every/each boy*, the existential quantifier *two (of the) boys* can grammatically combine with a collective predicate such as *formed a team*. This fact shows that the quantifier *two (of the) boys* is not distributive in lexicon, and more specifically, it should not be defined as existentially distributing over two atomic boys.

(68) a. Every/Each boy joined/*formed a team.

b. Two (of the) boys joined/formed a team.

The composition of *two of the boys* precedes as in (69). First, *of* combines with an entity denoted by the *the*-phrase and returns a set of subparts of this entity. Next, the numeral *two*, as a basic predicate restrictor, combines with a set of entities and returns a subset consisting of only the entities that have exactly two atomic (AT) subparts, as in (69c-d). Finally, a covert existential determiner $\varnothing_\exists$ combines with this set-denoting NumP and returns an existential generalized quantifier (Link 1987).

(69) *two of the boys*



Assume that the discourse domain has three boys *abc*:

a. $[\![\textit{the boys}]\!] = a \oplus b \oplus c$

b. $[\![\textit{of}]\!] = \lambda x_e.\{y \mid y \leq x\}$

c. $[\![\textit{two}]\!] = \lambda Q_{\langle e,t\rangle}.\{x \mid |\text{AT}(x)| = 2 \wedge x \in Q\}$

d. $[\![\textit{two of the boys}_@]\!] = \{a \oplus b, b \oplus c, a \oplus c\}$

e. $[\![\varnothing_\exists \textit{ two of the boys}_@]\!] = \lambda P_{\langle e,t\rangle}.\exists x[|\text{AT}(x)| = 2 \wedge \text{Bs}_@(x) \wedge P(x)]$
$= \lambda P_{\langle e,t\rangle}.\exists x \in \{a \oplus b, b \oplus c, a \oplus c\}[P(x)]$

Return to the composition of a matrix ∃2-question. In the following, $\textsf{2-Bs}_@$ abbreviates for the set of entities that are pluralities of two boys in the actual world. The same as in (65b), here the eligible minimal $\textsf{K}$ sets yielded by the application of the E-MIN-operator are all singleton sets, each of these sets consists of a proposition of the form $\ulcorner x$ *watched* $\mathbf{f}(x)\urcorner$ where x is the plurality of two boys, as in (70b). Hence, the derived reading is [−PL].

(70) Which movie did two of the boys watch? (QIQ-reading)

$[_{\text{CP}}\text{BeDom(wh-movie}_@)\ \lambda \mathbf{f}_{\langle e,e\rangle}\ [_2 \bigcap [_\gamma\ \text{E-MIN}\ \lambda\textsf{K}\ [_1\ \text{two-boys}_@\ \lambda x\ [\textsf{K}(\lambda w.x\text{-watch}_w\text{-}\mathbf{f}(x))]]]]]$

a. $[\![1]\!] = \exists x \in \textsf{2-Bs}_@[\textsf{K}(\lambda w.\textsf{W}_w(x,\mathbf{f}(x)))]$

b. $[\![\gamma]\!] = [\![\text{E-MIN}]\!](\lambda \textsf{K}.[\![1]\!])$
$\quad = \{\lambda w.\textsf{W}_w(x,\mathbf{f}(x))\}$, where $x = f_{\text{CH}}(\textsf{2-Bs}_@)$

c. $[\![2]\!] = \lambda w.\textsf{W}_w(x,\mathbf{f}(x))$, where $x = f_{\text{CH}}(\textsf{2-Bs}_@)$

d. $[\![\text{CP}]\!] = \lambda \mathbf{f}_{\langle e,e\rangle} : \text{Ran}(\mathbf{f}) \subseteq \textsf{M}_@.\lambda w[\textsf{W}_w(x,\mathbf{f}(x))]$, where $x = f_{\text{CH}}(\textsf{2-Bs}_@)$

In contrast to matrix ∃-questions, extensional embeddings of ∃-questions sometimes admit pair-list readings (Szabolcsi 1997a; Beghelli 1997; also see Appendix B). For example, the sentence (71) is felicitous even if each boy watched a different movie. I assume that this embedding sentence has the LF in (71a) and is interpreted as in (71b). In this LF, the existential indefinite moves over the embedding verb *know*, and its trace in the matrix clause is associated with a covert distributor EACH, which yields the 'EACH ≫ ι' reading.

(71) Susi knows [which movie two of the boys watched]. (∃2 ≫ EACH ≫ ι)

a. $[[\varnothing_\exists\ \text{two-boys}_@]\ \lambda x_e\ [[x\ \text{EACH}]\ \lambda y_e$ [Susi knows which movie y watched]]]

b. $\exists x[x \in \textsf{2-Bs}_@ \wedge \forall y \in \text{At}(x)[[\![\textit{Susi knows which movie y watched}]\!]]]$

In matrix ∃-questions, however, pair-list readings cannot be licensed by VP-*each*. In (72), the meaning contribution of the distributor *each* is just that the two chosen boys watched the movie separately, not that they watched possibly a different movie. The presented analysis explains the fact easily: to derive a pair-list reading as in ∀-questions, the quantificational predication condition has to be distributive. Such distributivity arises only if (i) the quantifier itself is distributive, or if (ii) an additional distributor appears between the quantifier and the null predication operator $\textsf{K}$. Condition (i) is easily seen in questions with an *each*/*every*-subject. Condition (ii) does not apply to English, because VP-*each* can only be interpreted within IP as in (72a), not as high as in (72b).

(72) Which movie did two of the boys **each** watch? (∃2 ≫ ~~EACH~~ ≫ ι)

a. ... $[\bigcap [\text{E-MIN}\ \lambda\textsf{K}\ [\text{two-boys}_@\ \lambda x_e\ [\textsf{K}\ [_{\text{IP}}\ \lambda w\ [[x\ \textbf{each}]\ \lambda y_e\ [y\ \text{watch}_w\ \mathbf{f}(y))]]]]]]]$

b. * ... $[\bigcap [\text{E-MIN}\ \lambda\textsf{K}\ [\text{two-boys}_@\ \lambda x_e\ [[x\ \textbf{each}]\ \lambda y_e\ [\textsf{K}\ [_{\text{IP}}\ \lambda w\ [y\ \text{watch}_w\ \mathbf{f}(y))]]]]]]]$

6.3.3. Composing NO-questions

Recall that negative quantifiers do not license QIQ-readings. For example, the NO-question (73) cannot be responded by silence. This question admits only individual readings and functional readings.

(73) Which movie did {no boy, none of the boys} watch? (✓Individual, ✓Functional, ✗QIQ)

a. *Hulk.*

b. The movie that his grandpa recommended.

c. # [Silence]

The proposed analysis easily explains the deviance of the QIQ-reading in a NO-question. The minimal set that contains no proposition of the form $\ulcorner$*boy-x watched* $\mathbf{f}(x)\urcorner$ is simply the empty set, whose conjunction is undefined. Hence, composing the NO-question (73) using the LF schema in Figure 3 yields a function that maps each input Skolem function to undefinedness. The main steps of the composition are given as follows:

(74) Which movie did no boy watch? (#QiQ-reading)

$[_{\text{CP}}$BEDOM(wh-movie$_@$) $\lambda \mathbf{f}_{\langle e,e\rangle}$ $[_2 \bigcap [_\gamma$ E-MIN $\lambda \mathsf{K}_{\langle st,t\rangle}$ $[_1$ no-boy$_@$ λx_e [$\mathsf{K}(\lambda w.x$-watch$_w$-$\mathbf{f}(x))$]]]]]

a. $[\![\textit{no boy}_@]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[\mathsf{boys}_@(x) \wedge P(x)]$

b. $[\![1]\!] = \neg\exists x \in \mathsf{B}_@[\mathsf{K}(\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)))]$

c. $[\![\gamma]\!] = [\![\text{E-MIN}]\!](\lambda \mathsf{K}.[\![1]\!]) = \varnothing$

d. $[\![2]\!]$ is undefined

6.3.4. Questions with a counting quantifier

It looks appealing and simple to extend the analysis in Sect. 6.3.3 for negative quantifiers to other decreasing quantifiers. For example, as seen in (75), decreasing quantifiers such as *at most two boys* and *less than three boys* also do not license QIQ-readings. The boy(s)-movie pair answer (75b) must be read in the same way as the individual answer (75a) except that the two boys are named explicitly, and the uniqueness inference triggered by *which movie* must be interpreted globally.

(75) Which movie did {at most two, less than three} boys watch?

\# 'For {at most two, less than three} boys x, [tell me] which unique movie did x watch?'

a. *Hulk*. (Intended: '*Hulk* is the only movie watched by {at most two, less than three} boys. The other movies were watched by more boys.')

b. Andy and Billy watched *Hulk*.

  i. ✓ Individual reading: '*Hulk* is the **only** movie watched by {at most two, less than three} boys, who are Andy and Billy. The other movies were watched by more boys.'

  ii. ✗ Choice reading: 'Andy and Billy are two boys who both watched **only** *Hulk*.'

In Xiang 2019a, following Hackl (2000), I decomposed a decreasing quantifier into a negative determiner *no* and a set-denoting restrictor, as in (76). With this decompositional analysis, the unavailability of QIQ-readings in (75) can be explained in the same way as in (74).

(76) a. $[\![\textit{at most two boys}_@]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[\#\text{AT}(x) > 2 \wedge \mathsf{Bs}_@(x) \wedge P(x)]$

b. $[\![\textit{less than three boys}_@]\!] = \lambda P_{\langle e,t\rangle}.\neg\exists x[\#\text{AT}(x) \geq 3 \wedge \mathsf{Bs}_@(x) \wedge P(x)]$

However, the questions in (77) do not admit QIQ/choice-readings either, despite that the quantifiers *at least two boys* and *exactly two boys* are not decreasing. The same as in (75), here the uniqueness inference triggered by the singular-marked *wh*-object has to be interpreted above the subject-quantifier. This fact shows that the unavailability of QIQ-readings in (75) and (77) has nothing to do with the monotonicity pattern of the subject-quantifier.

(77) Which movie did {at least, exactly} two boys watch? (✓Individual, ✓Functional, ✗QIQ)
'For {at least two, exactly two} boys x, [tell me] which unique movie did x watch?

In contrast to my old analysis, I now argue that the unavailability of QIQ-readings in (75) and (77) is due to a general syntactic constraint that counting quantifiers are scopally unproductive (Szabolcsi 1997b; Beghelli and Stowell 1997; among others). Beghelli and Stowell (1997) distinguish between the following four types of non-interrogative quantifiers and argue that they have different landing sites. In particular, counting quantifiers have very local scope and resist specific interpretations.

(78) **Types of non-interrogative quantifiers** (Beghelli and Stowell 1997)

a. Negative quantifiers: *no*-NP.
b. Universal-distributive quantifiers: *every/each*-NP
c. Grouping quantifiers: indefinites like *a*/*some*/*several*-NP, bare-numeral quantifiers (e.g., *one student*, *three students*), and *the*-phrases.
d. Counting quantifiers: decreasing quantifiers headed with determiners like *few*, *fewer than five*, and *at most six*; cardinality expressions with a modified numerals (e.g., *more than five*, *between six and nine*).

To derive the QIQ-reading of a question, the quantifier in this question must escape the IP and take scope above a null predicative operator K. Counting quantifiers cannot land at such a high position and thus do not license QIQ-readings.

6.4. Summary

To sum up the core analysis, I have argued that pair-list readings of multi-*wh* questions and QIQ-readings of questions with quantifiers are extensional functional readings — the object-*wh* leaves a complex functional trace, in which the argument index is bound by the subject-*wh*/quantifier. As generalized in (79a) and (79b), for both types of questions, the composition of the question nucleus precedes in three steps as described in (80).

(79) a. Which boy watched which movie? (Pair-list reading)
... wh-movie$_{@}$ $\lambda \mathbf{f}_{\langle e,e\rangle}$ [$_{\text{C}}$ $\bigcap$ $\lambda p_{\langle s,t\rangle}$ [$_{\text{B}}$ wh-boy$_{@}$ λx_e [[ID p] [$_{\text{A}}$ $\lambda w.x$ watched$_w$ $\mathbf{f}(x)$]]]]

b. Which movie did Det-boy watch? (QIQ-reading)
... wh-movie$_{@}$ $\lambda \mathbf{f}_{\langle e,e\rangle}$ [$_{\text{C}}$ $\bigcap$ E-MIN $\lambda \mathsf{K}_{\langle st,t\rangle}$ [$_{\text{B}}$ Det-boy$_{@}$ λx_e [K [$_{\text{A}}$ $\lambda w.x$ watched$_w$ $\mathbf{f}(x)$]]]]

(80) A. Indexations with the traces of the quantifiers and the *wh*-phrases yield an open sentence expressing functional dependency;
B. The subject-*wh*/quantifier quantifies-into an identity/predication condition of the functionality sentence;
C. Conjoining a set of propositions that have a dependency form yields a function graph.

Table 2 compares the nucleus denotations of four related pair-list multi-*wh* questions and QIQ-questions. In all questions, the asserted component of the nucleus denotation is the conjunction of a proposition set that describes the graph of the input Skolem function **f**. In the three questions with a non-interrogative subject-quantifier, the quantificational predication condition yielded at the B-node gives rise to an additional definedness condition which restricts the domain of the input Skolem

subject-type	Domain condition of $\mathbf{f}$	Function graph of $\mathbf{f}$	D-EXH	PL	CH
which boy		$\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}$	$-$	$+$	$-$
every/each boy	$\forall x \in \mathsf{B}_@[x \in \text{Dom}(\mathbf{f})]$	$\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x)) \mid \mathsf{B}_@(x)\}$	$+$	$+$	$-$
n of the boys	$\exists x \in n\text{-}\mathsf{Bs}_@[x \in \text{Dom}(\mathbf{f})]$	$\bigcap\{\lambda w.\mathsf{W}_w(x, \mathbf{f}(x))\}$ $(x \in n\text{-}\mathsf{Bs}_@)$	$-$	$-$	$+$
none of the boy	$\neg\exists x \in \mathsf{B}_@[x \in \text{Dom}(\mathbf{f})]$	$\bigcap \varnothing$	$-$	$-$	$-$

Table 2: Comparing the denotation of the question nucleus

function $\mathbf{f}$. In contrast, the multi-*wh* question does not have this condition and therefore is not subject to domain exhaustivity.

The QIQ-effect in questions with a non-interrogative quantifier is derived by extracting one of the minimal proposition sets that satisfy the quantificational predication condition yielded at the B-node. This analysis naturally explains the differences among the QIQ-questions with respect to the following three parameters:

- [±D-EXH]: As in the ∀-question, the yielded QIQ-reading presupposes **domain exhaustivity** if the quantificational predication condition is subject to a definedness condition that the input $\mathbf{f}$ is defined for every element in the quantification domain of the subject.
- [±PL]: As in the ∀-question, with other conditions being equal, the yielded QIQ-reading admits **pair-list** answers only if there is a non-singleton set of propositions that minimally satisfies the quantificational predication condition yielded at B.
- [±CH]: As in the ∃-question, with other conditions being equal, the yielded QIQ-reading has a **choice** flavor if there are multiple minimal proposition sets satisfying the quantificational predication condition yielded at B.

In addition to questions with a universal quantifier or an existential indefinite, this section has also explained why in many cases QIQ-readings are unavailable. In questions with a negative quantifier (e.g., *no boy*, *none of the boys*), QIQ-readings are deviant because the only minimal proposition set that satisfies a negative-quantificational predication condition is the empty set. In questions with a counting quantifier (e.g., *exactly two boys*, *two or more (of the) boys*), the LF used for deriving QIQ-readings is unavailable because counting quantifiers are unproductive in scoping.

7. Quantificational variability effects

As seen in Sect. 4.1.2, defining pair-list questions as sets of conjunctive propositions, the functionality-based approach of Dayal (1996, 2017) cannot account for the QV effects in embeddings of pair-list questions. Dayal defines simplex and pair-list questions uniformly as sets of propositions. In the case of embedding a simplex *wh*-question, the most natural way for Dayal to define a QV inference is to let the matrix quantification adverbial quantifies over a set of atomic propositions, as in (81).

(81) Jill mostly knows [which students left].
$\rightsquigarrow$ *For most p: p is a true proposition of the form ⌜student-x left⌝, Jill knows p.*

This proposition-based definition of QV inferences, however, is not feasible for embeddings of pair-list questions if pair-list questions are defined as conjunctive propositions (Lahiri 2002). For example, in case that three relevant boys $b_1 b_2 b_3$ watched movies $m_1 m_2 m_3$, respectively, the strongest true propositional answer of the embedded pair-list question in (82) is $\lambda w.\mathsf{W}_w(m_1, b_1) \wedge \mathsf{W}_w(m_2, b_2) \wedge \mathsf{W}_w(m_3, b_3)$, and the embedding sentence is true only if Jill knows at least two of the three atomic conjuncts, as in (82a); however, these conjuncts cannot be semantically retrieved out of their conjunction. In contrast, family-of-questions approaches such as Fox 2012a,b can derive this QV inference by defining the quantification domain as a set of sub-questions as in (82b).

(82) Jill mostly knows $[_{\text{PAIR-LIST}} \left\{ \begin{array}{l} \text{which movie every boy watched} \\ \text{which boy watched which movie} \end{array} \right\}]$.

a. ⇝ *For most p s.t. p is a true proposition of the form ⌜boy-x watched movie-y⌝, Jill knows p.*

b. ⇝ *For most Q s.t. Q is a question the form ⌜which movie boy-x watched⌝, Jill knows Q.*

c. ⇝ *For most* $\langle x, y\rangle$ *s.t. x is an atomic boy, y is an atomic movie, and x watched y, Jill knows that x watched y.*

Although this paper does not pursue a family-of-questions approach, the assumed hybrid categorial approach to question composition unlocks the option (82c), where the quantification domain of *mostly* is a set of atomic pairs. In my proposal, a pair-list question denotes a topical property that maps each input Skolem function to a conjunctive proposition. From this topical property, we can extract the Skolem function that yields the strongest true answer to this question and define the quantification domain of *mostly* as a set of atomic subparts of this Skolem function. For example in (84), the strongest true answer is the Skolem function (84a), and the atomic boy-watch-movie pairs are as in (84b).

(83) a. A function f is **atomic** if and only if $\bigoplus \text{Dom}(f')$ is atomic.

b. $\textsc{At}(\mathbf{f}) = \{f' \mid f' \subseteq \mathbf{f} \text{ and } f' \text{ is atomic}\}$

(84) Which boy watched which movie?/ Which movie did every boy watch?
(Context: *The discourse domain includes three boys $b_1 b_2 b_3$ and three movies $m_1 m_2 m_3$. In a world w, b_1 watched only m_1, b_2 watched only m_2, and b_3 watched only m_3.*)

a. $\textsc{Ans}^S(w)([\![Q]\!]) = \begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$

b. $\textsc{At}(\textsc{Ans}^S(w)([\![Q]\!])) = \left\{ \begin{array}{c} [b_1 \to m_1] \\ [b_2 \to m_2] \\ [b_3 \to m_3] \end{array} \right\}$

Xiang 2020 provides two ways to define a QV inference based on short answers. Ignoring the complications needed for accounting for mention-some readings, I schematize these two definitions as in (85a-b).[15] (For a compositional derivation, see Cremers 2018.) In both definitions, the quantification domain of the matrix adverbial *mostly* is a set of atomic entities or a set of atomic Skolem functions.

(85) The **QV inference** of 'Jill mostly knows Q'

a. $\lambda w.\textsc{Most}\ x[x \in \textsc{At}(\textsc{Ans}^S(w)([\![Q]\!]))][\mathsf{know}_w(j, [\![Q]\!](x)]$
(For most x such that x is an atomic subpart of the strongest true short answer to Q, Jill knows the inference $[\![Q]\!](x)$.)

b. $\lambda w.\textsc{Most}\ x[x \in \textsc{At}(\textsc{Ans}^S(w)([\![Q]\!]))][\mathsf{know}_w(j, \lambda w'.x \leq \textsc{Ans}^S(w')([\![Q]\!]))]$
(For most x such that x is an atomic subpart of the strongest true short answer to Q, Jill knows that x is a subpart of the strongest true short answer to Q.)

[15] Xiang (2020) considers also mention-some readings of questions, where a question can have multiple complete true answers. Once mention-some reading is concerned, $\textsc{Ans}^S(w)(Q)$ needs to be defined as a set of entities/functions, not one single entity/function.

In (85a), the scope of the adverbial *mostly* says that Jill knows an atomic proposition, which is derived by applying the topical property of the embedded question to an entity/function x, where x is an atomic subpart of the strongest true answer to the embedded question. This definition works for embeddings of multi-*wh* questions, but not for embeddings of ∀-question: the topical property of the ∀-question *which movie every boy watched* is only defined for Skolem functions that are defined for every boy, not for atomic Skolem functions such as $[b_1 \to m_1]$.

Alternatively, in (85b), the scope of *mostly* says that Jill knows a sub-divisive inference, which is semantically equivalent to that Jill correctly identifies most of the boy-watched-movie pairs. In the context described in (84), this sub-divisive inference is true if and only if in every world w' such that w' is compatible with Jill's belief, the strongest true short answer to the embedded ∀-question in w' is among the seven Skolem functions in Figure 5. This figure illustrates a partition of possible worlds grouped based on which movie each of the boys watched. The world w described in (84) belongs to the middle cell. In the other cells, correspondences conflicting with w are colored in gray. It is straightforward to see that the union of the seven cells is equivalent to the following inference: 'for most of the pairs $\langle b, m\rangle$ in $\{\langle b_1, m_1\rangle, \langle b_2, m_2\rangle, \langle b_3, m_3\rangle\}$, b watched m.' Knowing this inference simply means correctly identifying most of the three boy-watched-movie pairs.

	$\begin{bmatrix} b_1 \to m_2 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_3 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$
$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_1 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_3 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_3 \\ b_3 \to m_3 \end{bmatrix}$
$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_1 \end{bmatrix}$	$\begin{bmatrix} b_1 \to m_1 \\ b_2 \to m_2 \\ b_3 \to m_2 \end{bmatrix}$	

Figure 5: Illustration of the sub-divisive inference in the quantification scope of (85b)

8. Conclusions

In this paper, I pointed out that pair-list ∀-questions and their multi-*wh* counterparts are semantically different — only the former are subject to domain exhaustivity. This difference suggests that the structure of composition of a pair-list ∀-question must be distinct from that of its multi-*wh* counterpart. Furthermore, the uniform syntactic constraints on distributing QIQ-readings show that QIQ-readings of matrix questions should be derived uniformly.

Influential accounts such as Dayal 1996, 2017 and Fox 2012a,b have not noticed the contrast between ∀- and multi-*wh* questions with respect to domain exhaustivity. These accounts treat pair-list questions uniform and compose these questions either with the same LF or with different LFs that yield the same root denotation. In addition, to explain why only subject *every*/*each*-phrases license pair-list readings, these accounts derive pair-list readings in a way that crashes in questions with a non-universal quantifier. In consequence, they overly predict domain exhaustivity effects for multi-*wh* questions and fail to extend to ∃-questions with choice readings.

This paper provided a novel analysis to compose complex questions. This analysis has three core parts. **First**, in line with functionality-based approaches, I proposed that QIQ-questions and pair-list multi-*wh* questions both involve *wh*-dependencies — the subject-*wh*/quantifier binds the argument variable of the functional trace of the *wh*-object. In particular, in a pair-list multi-*wh* question, the subject-*wh* quantifies into an identity condition with respect to this *wh*-dependency relation; in a QɪQ-question, the subject-quantifier quantifies-into a predication relation with respect to this dependency. The subtle differences between the two quantifying-in operations are responsible of the contrast between ∀- and multi-*wh* questions with respect to domain exhaustivity. **Second**, for questions with quantifiers in specific, inspired by Fox (2012b), I assumed that the seeming QɪQ-effect is derived by extracting one of the minimal proposition sets that satisfy the quantificational predication condition. This analysis naturally predicts which questions admit QɪQ-readings and whether their QɪQ-readings are subject to domain exhaustivity, admit pair-list answers, and have a choice flavor. **Finally**, adopting the hybrid categorial approach to compose questions, the presented analysis also overcame the difficulty with the functionality-based analysis of Dayal 1996 in accounting for QV effects in embeddings of pair-list questions.

Appendix A. The partition-based approach

Section 3 has mentioned that the following LF, repeated from (17), suffers type-mismatch for most frameworks of question semantics:

(86) Which movie did Det-boy watch?
*[Det-boy λx_e [which movie did x watch]]

Partition Semantics is an exception to this type-mismatch problem. Groenendijk and Stokhof (1984: chapter 3) first analyze the pair-list ∀-question (87) as a partition of possible worlds grouped in terms of which boy watched which movie. In the derivation of this denotation, the quantifier *every boy* quantifies into an identify condition (of type t), which says that x watched the same movies in w and in w'.

(87) Which movie did every boy watch?
$\lambda w \lambda w'.\forall x[\mathsf{B}_@(x) \rightarrow \{y \mid \mathsf{M}_@(y) \wedge \mathsf{W}_w(x,y)\} = \{y \mid \mathsf{M}_@(y) \wedge \mathsf{W}_{w'}(x,y)\}]$
(w and w' are in the same partition cell if and only if for every boy x, x watched the same movies in w and in w'.)

However, Groenendijk and Stokhof themselves are not satisfied with this account since it does not extend to questions with a non-universal quantifier. For example, the predicted meaning for the corresponding ∃-question (88) is not a partition (see also Krifka 2001). Thus, Groenendijk and Stokhof ultimately pursues another family-of-question approach using witness sets (see footnote 10).

(88) Which movie did one of the boys watch?
$\lambda w \lambda w'.\exists x[\mathsf{B}_@(x) \wedge \{y \mid \mathsf{M}_@(y) \wedge \mathsf{W}_w(x,y)\} = \{y \mid \mathsf{M}_@(y) \wedge \mathsf{W}_{w'}(x,y)\}]$
(w and w' are in the same partition cell if and only if for one of the boys x, x watched the same movies in w and in w'.)

For illustration, consider a discourse with two boys *ab* and two movies $m_1 m_2$. The four worlds vary by which boy watched which movie. $w_1 w_2 w_3$ are grouped in one cell C_1: a watched the same movie in w_1 and w_2 (and b watched the same movie in w_1 and w_3). Likewise, $w_2 w_3 w_4$ are in one cell C_2: b watched the same movie in w_2 and w_4. In addition, C_1 and C_2 are distinct cells because neither boy watched the same movie in w_1 and w_4. The world grouping in Fig. 6 is clearly not a partition: C_1 and C_2 are overlapped, both containing w_2 and w_3. Moreover, from this world grouping, we cannot identify which movie any of the boys watched. For example, if w_1 is the actual world, then C_1 is the cell which the actual world belongs to; however, based on C_1, we cannot decide on whether a watched m_1 (as in w_1 and w_2) or he watched m_2 (as in w_3).

$$C_1 : \begin{array}{|l|}\hline w_1 : \{\langle a, m_1\rangle, \langle b, m_2\rangle\} \\ w_2 : \{\langle a, m_1\rangle, \langle b, m_1\rangle\} \\ w_3 : \{\langle a, m_2\rangle, \langle b, m_2\rangle\} \\ \hline w_4 : \{\langle a, m_2\rangle, \langle b, m_1\rangle\} \\ \hline \end{array} \qquad C_2 : \begin{array}{|l|}\hline w_1 : \{\langle a, m_1\rangle, \langle b, m_2\rangle\} \\ \hline w_2 : \{\langle a, m_1\rangle, \langle b, m_1\rangle\} \\ w_3 : \{\langle a, m_2\rangle, \langle b, m_2\rangle\} \\ w_4 : \{\langle a, m_2\rangle, \langle b, m_1\rangle\} \\ \hline \end{array} \qquad ......$$

Figure 6: World grouping yielded by (88)

In addition, this analysis inherits the theory-internal problems with Partition Semantics. For instance, Partition Semantics cannot explain the uniqueness effects of singular-marked *wh*-questions (Xiang 2020); likewise, the partition-based account cannot explain the point-wise uniqueness effects in pair-list ∀-questions.

Appendix B. The question-embedding approach

Another intuitive and framework-independent way to solve the type-mismatch in quantifying-into questions is to reduce matrix questions into question-embeddings and let the quantifier take scope over a covert question-embedding predicate (Karttunen 1977; Krifka 2001). The LF assumed by Karttunen (1977) is given in (89). Basically, whatever the embedded question denotes, the question-embedding is a t-type expression which can be quantified into.

(89) Which movie did Det-boy(s) watch?
[Det-boy(s) λx_e [I-ASK-YOU [which movie did x watch]]]

This analysis relies on the quantifier in the embedded question taking scope over an intensional embedding predicate (e.g., *ask*, *wonder*). However, drawing on the limited distribution of pair-list readings in matrix questions and intensional question-embeddings, I argue that this scoping pattern is not available.[16]

As seen in Sect. 3 and explained in Sect. 6.3, only *every*/*each*-phrases may license pair-list readings of matrix questions. As for question-embeddings, Szabolcsi (1997a) observes a contrast between intensional complements and extensional complements.[17] In particular, in embeddings with an extensional predicate (e.g., *know*, *find out*), numeral-modified indefinites such as *two of the boys* may also license a pair-list reading. For example, in a context assuming point-wise uniqueness from boys to movies, the sentences (90a-b) are felicitous and can be read with the following scopal pattern: '$\exists 2 \gg$ EACH $\gg$ V $\gg \iota$' where 'V' stands for the embedding predicate. As I have argued in Sect. 6.3.2, this reading can be derived from the LF in (91): the existential indefinite takes wide scope relative to the embedding predicate *know*, and its closest trace in the matrix clause is associated with a covert

[16]Krifka (2001) assumes the structure in (i) where the quantifier scopes over a speech act operator QUEST. This analysis exempts from the over-generation problem since Krifka assumes that speech acts cannot be disjoined. However, it also leaves the choice readings of ∃-questions unexplained.

(i) Which movie did every boy watch?
[every-boy λx_e [QUEST [which movie did x watch]]]

[17]The intension-vs-extension qualification comes from Groenendijk and Stokhof 1984. In later works starting from Lahiri 2002, this division is re-labeled as 'rogative'-vs-'responsive'. Rogative predicates admit only interrogative complements, while responsive predicates admit also declarative complements.

distributor EACH (see also (71)).[18]

(90) Susi knew that each boy watched a different movie. In addition, ...
  a. Susi **knew** which movie each/two of the boys watched.
  b. Susi **found out** which movie each/two of the boys watched.

(91) Susi V-ed which movie two of the boys watched.
[[two-of-the-boys λx [[EACH x] λy [Susi V-ed which movie y watched]]]

However, embeddings with an intensional predicate behave the same as matrix questions — only *every*/*each*-phrases license pair-list readings in these embeddings. For example, in (92a-b), the uniqueness inference triggered by the singular-marked object *which movie* must be interpreted between the embedding predicate and the quantifier: ASK $\gg \iota \gg \exists 2$. The lack of pair-list readings shows that the sentences (92a-b) cannot have the LF the (91). As Szabolcsi argues, a natural explanation to the unavailability of this LF would be that intensional verbs create weak islands, which prevent the quantifiers in the embedded questions from taking wide scope. If this explanation is on the right track, the embedding structure (89) should not be feasible either.

(92) Susi knew that every boy watched a different movie. ...
  a. Susi **wondered** which movie each/#two of the boys watched.
  b. Susi **asked me** which movie each/#two of the boys watched.

In short, QIQ-effects in matrix questions should not be treated as quantification into question-embeddings. Even if matrix questions could be viewed as embeddings with a covert *ask*, the quantifier in an embedded question cannot scope over this *ask*.

Acknowledgement [To be added ...]

References

Barwise, Jon, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.

Beck, Sigrid, and Yael Sharvit. 2002. Pluralities of questions. *Journal of Semantics* 19:105–157.

Beghelli, Filippo. 1997. The syntax of distributivity and pair-list readings. In *Ways of scope taking*, ed. Anna Szabolcsi, 349–408. Dordrecht: Kluwer Academic Publishers.

Beghelli, Filippo, and Tim Stowell. 1997. Distributivity and negation: The syntax of *each* and *every*. In *Ways of scope taking*, ed. Anna Szabolcsi, 71–107. Dordrecht: Kluwer Academic Publishers.

Berman, Stephen Robert. 1991. On the semantics and logical form of *wh*-clauses. Doctoral Dissertation, University of Massachusetts at Amherst.

[18] Instead of using covert movement, Szabolcsi (1997a) derives the wide scope reading by type-lifting the interrogative complements of extensional predicates. Combining the type-lifted question-denotation (i) with an embedding predicate P yields a wide scope reading of the quantifier π relative to P. Further, Szabolcsi argues that *wonder*-type predicates cannot select for lifted questions and hence that quantifiers in intensional complements cannot take wide scope.

(i) Complement of *find out*-type predicates: $\lambda P.\pi(\lambda x.P(\text{which } y[x \text{ watched } y]))$

Caponigro, Ivano, and Anamaria Fălăuş. To appear. Unveiling multiple *wh*-free relative clauses and their functional *wh*-words. *Semantics and Pragmatics* .

Chierchia, Gennaro. 1991. Functional WH and weak crossover. In *The Proceedings of the West Coast Conference on Formal Linguistics 10*, ed. Dawn Bates, 75–90.

Chierchia, Gennaro. 1993. Questions with quantifiers. *Natural Language Semantics* 1:181–234.

Ciardelli, Ivano, and Floris Roelofsen. 2018. An inquisitive perspective on modals and quantifiers. *Annual Review of Linguistics* 4:129–149.

Cremers, Alexandre. 2016. On the semantics of embedded questions. Doctoral Dissertation, École normale supérieure, Paris.

Cremers, Alexandre. 2018. Plurality effects in an exhaustification-based theory of embedded questions. *Natural Language Semantics* 26:193–251.

Dayal, Veneeta. 1996. *Locality in wh-quantification: Questions and relative clauses in Hindi*. Dordrecht: Kluwer.

Dayal, Veneeta. 2002. Single-pair versus multiple-pair answers: Wh-in-situ and scope. *Linguistic Inquiry* 33:512–520.

Dayal, Veneeta. 2016. List answers through higher order questions. Colloquium talk at MIT, February 2016.

Dayal, Veneeta. 2017. *Questions*. Oxford: Oxford University Press.

Engdahl, Elisabet. 1980. The syntax and semantics of questions in Swedish. Doctoral Dissertation, University of Massachusetts at Amherst.

Engdahl, Elisabet. 1986. *Constituent questions*. Dordrecht: Reidel.

Fox, Danny. 2012a. Multiple wh-questions: Uniqueness, pair-list and second order questions. Class notes for MIT seminars.

Fox, Danny. 2012b. Pair-list with universal quantifiers. Class notes for MIT seminars.

Fox, Danny. 2013. Mention-some readings of questions. Class notes for MIT seminars.

Gallin, Daniel. 1975. *Intensional and higher-order modal logic*. Amsterdam: North-Holland Publishing Company.

Groenendijk, Jeroen, and Martin Stokhof. 1984. On the semantics of questions and the pragmatics of answers. *Varieties of formal semantics* 3:143–170.

Hackl, Martin. 2000. Comparative quantifiers. Doctoral Dissertation, Massachusetts Institute of Technology.

Hagstrom, Paul Alan. 1998. Decomposing questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Heim, Irene. 1995. Notes on questions. Class notes for MIT seminars.

Heim, Irene. 2012. Notes on questions. MIT class notes for Advanced Semantics.

Jacobson, Pauline. 1994. Binding connectivity in copular sentences. In *Proceedings of Semantics and Linguistic Theory*, ed. M. Harvey and L. Santelmann, volume 4, 161–178. URL `https://doi.org/10.3765/salt.v4i0.2456`.

Jacobson, Pauline. 1999. Towards a variable-free semantics. *Linguistics and Philosophy* 22:117–185.

Kang, Jungmin. 2012. Why short-form functional reading answers are not possible in multiple *wh*-questions. In *University of Pennsylvania Working Papers in Linguistics*, volume 18. URL https://repository.upenn.edu/pwpl/vol18/iss1/15.

Karttunen, Lauri. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1:3–44.

Kotek, Hadas. 2014. Composing questions. Doctoral Dissertation, Massachusetts Institute of Technology.

Kotek, Hadas. 2019. *Composing questions*, volume 80. Cambridge, MA: MIT Press.

Krifka, Manfred. 1991. Definite noun-phrases are not quantifiers. *Linguistic Inquiry* 23:156–163.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9:1–40.

Lahiri, Utpal. 1991. Embedded interrogatives and predicates that embed them. Doctoral Dissertation, Massachusetts Institute of Technology.

Lahiri, Utpal. 2002. *Questions and answers in embedded contexts*. Oxford: Oxford University Press.

Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In *Meaning, use, and interpretation of language*, ed. Christoph Schwarze Rainer Bäuerle and Arnim von Stechow, 302–323. De Gruyter.

Link, Godehard. 1987. Generalized quantifiers and plurals. In *Generalized quantifiers*, ed. P. Gärdenfors, 151–180. Dordrecht: Reidel.

May, Robert. 1985. *Logical form: Its structure and derivation*, volume 12. Cambridge, MA: MIT press.

May, Robert. 1988. Ambiguities of quantification and *wh*: A reply to Williams. *Linguistic Inquiry* 19:118–135.

Moltmann, Friederike, and Anna Szabolcsi. 1994. Scope interactions with pair-list quantifiers. In *Proceedings of the North East Linguistic Society 24*, ed. Mercé González, volume 2, 381–395. GLSA Publications, University of Massachusetts, Amherst.

Nicolae, Andreea Cristina. 2013. Any questions? Polarity as a window into the structure of questions. Doctoral Dissertation, Harvard University.

Pafel, Jürgen. 1999. Interrogative quantifiers within scope. *Linguistics and Philosophy* 22:255–310.

Partee, Barbara H. 1986. Noun phrase interpretation and type-shifting principles. In *Studies in discourse representation theory and the theory of generalized quantifiers*, ed. J. Groenendijk, D. de Jongh, and M. Stokhof, 357–381. Foris, Dordrecht.

Preuss, Susanne Marie-Luise. 2001. Issues in the semantics of questions with quantifiers. Doctoral Dissertation, Rutgers University.

Sharvit, Yael, and Jungmin Kang. 2017. Fragment functional answers. In *Proceedings of Semantics and Linguistic Theory*, ed. Mary Moroney, Carol-Rose Little, Jacob Collard, and Dan Burgdorf, volume 26, 1099–1118. URL https://doi.org/10.3765/salt.v26i0.3792.

Sharvy, Richard. 1980. A more general theory of definite descriptions. *The Philosophical Review* 89:607–624.

Srivastav, Veneeta. 1991. WH dependencies in Hindi and the theory of grammar. Doctoral Dissertation, Cornell University.

Szabolcsi, Anna. 1997a. Quantifiers in pair-list readings. In *Ways of scope taking*, ed. Anna Szabolcsi. Dordrecht: Kluwer Academic Publishers.

Szabolcsi, Anna. 1997b. Strategies for scope taking. In *Ways of scope taking*, ed. Anna Szabolcsi, 109–154. Dordrecht: Kluwer Academic Publishers.

Williams, Edwin Samuel. 1994. *Thematic structure in syntax*. 23. Cambridge, MA: MIT Press.

Winter, Yoad. 2001. *Flexibility principles in boolean semantics*. Cambridge, MA: MIT Press.

Xiang, Yimei. 2016. Interpreting questions with non-exhaustive answers. Doctoral Dissertation, Harvard University.

Xiang, Yimei. 2019a. Getting quantifying-into questions uniformly: Functionality, domain exhaustivity, and quantificational variability. In *Proceedings of Semantics and Linguistic Theory*, ed. Katherine Blake, Forrest Davis, Kaelyn Lamp, and Joseph Rhyne, volume 29, 160–179. URL https://doi.org/10.3765/salt.v29i0.4610.

Xiang, Yimei. 2019b. A variable-free approach to *wh*-dependency. Presentation at the Workshop on Complex wh-questions, Nantes, November 2019.

Xiang, Yimei. 2020. A hybrid categorial approach to question composition. *Linguistics and Philosophy* .